Geophysical Journal International

Geophys. J. Int. (2013) doi: 10.1093/gji/ggt284

Geophysical Journal International Advance Access published September 3, 2013

A Bayesian approach to infer radial models of temperature and anisotropy in the transition zone from surface wave dispersion curves

M. Drilleau,[1] É. Beucler,[1] A. Mocquet,[1] O. Verhoeven,[1] G. Moebs,[2] G. Burgos,[3] J.-P. Montagner[3] and P. Vacher[1]

[1] *Université de Nantes, Laboratoire de Planétologie et de Géodynamique, UMR-CNRS* 6112*, Nantes, France. E-mail: melanie.drilleau@univ-nantes.fr*
[2] *Université de Nantes, Laboratoire de Mathématiques Jean Leray, UMR-CNRS* 6629*, Nantes, France*
[3] *Institut de Physique du Globe de Paris, Jussieu, Laboratoire de Sismologie, UMR-CNRS* 7154*, Paris, France*

Accepted 2013 July 15. Received 2013 July 12; in original form 2013 April 5

SUMMARY

Mineralogical transformations and material transfers within the Earth's mantle make the 350–1000 km depth range (referred here as the mantle transition zone) highly heterogeneous and anisotropic. Most of the 3-D global tomographic models are anchored on small perturbations from 1-D models such as PREM, and are secondly interpreted in terms of temperature and composition distributions. However, the degree of heterogeneity in the transition zone can be strong enough so that the concept of a 1-D reference seismic model must be addressed. To avoid the use of any seismic reference model, we present in this paper a Markov chain Monte Carlo algorithm to directly interpret surface wave dispersion curves in terms of temperature and radial anisotropy distributions, here considering a given composition of the mantle. These interpretations are based on laboratory measurements of elastic moduli and Birch–Murnaghan equation of state. An originality of the algorithm is its ability to explore both smoothly varying models and first-order discontinuities, using C1-Bézier curves, which interpolate the randomly chosen values for depth, temperature and radial anisotropy. This parametrization is able to generate a self-adapting parameter space exploration while reducing the computing time. Thanks to a Bayesian exploration, the probability distributions on temperature and anisotropy are governed by uncertainties on the data set. The method is applied to both synthetic data and real dispersion curves. Though surface wave data are weakly sensitive to the sharpness of the of the mid-mantle seismic discontinuities, the interpretation of the temperature distribution is highly related to the chosen composition and to the modelling of mineralogical phase transformations. Surface wave measurements along the Vanuatu–California path suggest a strong anisotropy above 400 km depth, which decreases below, and a monotonous temperature distribution between 350 and 1000 km depth.

Key words: Inverse theory; Probability distributions; Equation of state; Phase transitions; Surface waves and free oscillations; Seismic anisotropy.

1 INTRODUCTION

Seismic tomography is one of the most powerful ways to provide information on the internal structure and dynamics of the Earth's deep interior. Seismic images reveal features at both lateral and radial resolutions which are continuously being improved. Various approaches for solving the inverse problem of describing the Earth's deep structure using seismological data sets have been implemented (e.g. Montagner & Tanimoto 1990; van der Hilst *et al.* 1997; Fukao *et al.* 2001; Romanowicz 2003). Although-high resolution earth's models are more and more numerous, a quantitative comparison is difficult due to the amount of various techniques providing different kind of uncertainties.

Recent 3-D seismic models (Zhu *et al.* 2012) are derived from sophisticated inverse theories and use accurate numerical methods for seismic wave propagation. While they improve the lateral resolution of the smooth 3-D starting models (Ritsema *et al.* 1999; Mégnin & Romanowicz 2000), they are intrinsically rooted to first-order perturbation theory from a laterally homogeneous reference models such as PREM (Dziewonski & Anderson 1981) or AK135 (Kennett *et al.* 1995). However, from a geodynamical point of view, in regions such as the transition zone, the degree of heterogeneity induced by mineralogical transformations, convective motions, upwelling and downwelling materials, and anisotropy, is strong enough so that the concept of a 1-D reference seismic model might be addressed.

Among the most recent studies on tomography, for example, Boschi & Dziewonski (2000), Trampert & Woodhouse (2003), Panning & Romanowicz (2006), Beghein *et al.* (2008), Kustowski *et al.* (2008) and Visser *et al.* (2008a,b), indicate radial anisotropic variations throughout the mantle, they only agree at long wavelengths (Kustowski *et al.* 2008; Panning *et al.* 2010), even when higher mode surface waves are used to investigate the transition zone depth range, hereafter defined between 350 and 1000 km depth range. These large uncertainties may be explained by the different inverse and regularization methods, the choice of the parametrization and the data set.

When going beyond the tomographic methods and addressing the question of the interpretation of the seismological information in terms of mantle temperature and composition, two approaches may be followed: by using either previously derived seismic models (Vacher *et al.* 1996, 1998; Goes *et al.* 2000; Trampert *et al.* 2004; Cammarano & Romanowicz 2007) or directly the observed traveltimes, surface wave velocities or eigenfrequencies (Cobden *et al.* 2008, 2009; Cammarano *et al.* 2009; Khan *et al.* 2009). One of the main advantages of the latter approach is to directly connect the posterior uncertainties on the geodynamical parameters to the variances on observations. However, the non-linear nature of the problem requires to use time-consuming Monte Carlo-like inverse methods. This difficulty has previously been handled by testing a limited amount of selected models (Cammarano *et al.* 2009), or by exploring tightly constrained parameter spaces (Khan *et al.* 2009, 2011; Khan & Shankland 2012).

In this paper, we choose to address these questions by inferring the radial temperature field and anisotropy distribution for a given mineralogy from surface wave data. This allows to relax the priors and to sample a large range of possible parameter values.

Bayesian approaches allow to go beyond the classical computation of the unique best-fitting model by providing a quantitative probabilistic measure of the model resolution, uncertainties and non-unicity (e.g. Mosegaard & Tarantola 1995). Though these methods such as the Markov chain Monte Carlo (McMC) are popular in geophysics, their use in global seismological studies and interpretation of seismological models in terms of mantle temperature and composition is recent. This is mainly due to the large amount of inverted parameters and to the huge required computing time (e.g. Verhoeven *et al.* 2009; Khan *et al.* 2009, 2011; Hauser *et al.* 2011; Bodin *et al.* 2012; Mosca *et al.* 2012). In a Bayesian framework, the known prior information on the parameters is combined with the observed data to generate the *a posteriori* distribution of the model parameters. McMC methods perform a non-linear guided search by sampling the parameter space according to the posterior probabilities. One important advantage of this method is the complete independence from the choice of the starting model.

In this study, we develop a non-linear inverse approach to directly interpret Love and Rayleigh surface waves dispersion curves of fundamental and higher modes, in terms of temperature and radial anisotropy distributions. The calculations are performed on the basis of a given mantle composition. Though seismic velocity variations are from both compositional and thermal origins, we tested that it is hardly possible to decorrelate these two effects only on the basis of surface wave seismic data. For example, a given decrease of the seismic velocities in the transition zone can be explained either by high temperatures, and/or by increasing the garnet content. Our approach shares in common some features of the forward problem addressed by Cammarano *et al.* (2009), who tested against the data thousands of temperature profiles for three given composition of the mantle; and a non-linear inverse method close to Khan *et al.* (2009, 2011), using an original parametrization and wider priors. Starting from random realizations of the temperature field and anisotropy distribution, the seismic velocity profiles are computed, using a Birch–Murnaghan equation of state. The synthetic dispersion curves are then obtained by normal-mode summation and compared to the data through an McMC method. An originality of the algorithm is its high flexibility during the exploration of the parameter space. The parametrization is defined by C1-Bézier curves, which interpolate the randomly sampled parameter values and generate a self-adapting parameter space exploration.

The forward problem is described in Section 2, while the Bayesian inversion is presented in Section 3. In Section 4, we detail the practical and numerical implementation of this procedure. Synthetic test results are exhibited in Section 5 and we illustrate the efficiency and the resolving power of the method by inverting phase velocities along the Vanuatu–California path, in terms of temperature and anisotropy of the transition zone.

2 FORWARD PROBLEM

This section describes the parametrization and modelling hypothesis, used to compute synthetic data for given temperature and anisotropy distributions, and the corresponding radial seismic models. To generate a self-adapting model space, with respect to the resolution power of the data and to slightly reduce the computing time, polynomial Bézier curves are chosen for the parametrization. This choice enables us to remove the effects of a regular depth-discretization, while allowing to generate a huge set of radial models.

2.1 Physical parameters

2.1.1 Isotropic parameters

A standard procedure to compute the density and elastic *moduli*, at a given depth, relies in extrapolating their values from standard temperature and pressure (STP) conditions, using thermodynamic laws (e.g. Jackson 1998; Stixrude & Lithgow-Bertelloni 2010). Laboratory experiments provide these STP values, their derivatives with respect to the temperature and pressure and the thermodynamic parameters from which they can be derived. The elastic behaviour of mantle materials may be computed starting either from their chemical or mineralogical compositions.

Here, the depth-dependent values of the density and the elastic *moduli* are first computed for each mineral using the approach described in Vacher *et al.* (1998), and the laboratory experiment values compiled by Cammarano *et al.* (2003) and Verhoeven *et al.* (2005). This method relies on a Birch–Murnaghan equation of state, up to third-order in strain, associated with a Grüneisen correction for the temperature. The elastic properties of the mineralogical assemblages are computed using the Hashin–Shtrikman estimate (Hashin & Shtrikman 1963; Watt *et al.* 1976). The functions needed to compute elastic parameters are strongly non-linear, that is why a Bayesian formulation is more appropriate than a deterministic (e.g. least-square) procedure.

While the temperature and pressure increase, the mantle minerals undergo mineralogical transformations, including phase transitions. The volume fractions of the mineral phases are computed using phase diagrams evaluated from laboratory measurements. Here, two independent subsystems are considered: the phase diagram of olivine and its high pressure phases, and a phase diagram for all other

components. Subsystems are computed using a Gibbs free-energy minimization method based on experimentally constrained phase boundaries (Ita & Stixrude 1992). The details of the phase diagrams can be found in Vacher *et al.* (1998). They incorporate changes in Fe/Mg ratios in the minerals with depth. The major considered mantle minerals are: olivine, wadsleyite, ringwoodite, Mg-perovskite, Mg-wüstite, clinopyroxene, orthopyroxene, Ca-pyroxene, ilmenite and majorite.

The corresponding radial density and seismic velocity models are then used to compute surface wave velocity dispersion curves, using the MINEOS package from CIG,[1] based on the pioneer work of Gilbert & Dziewonski (1975) updated by Woodhouse (1988) and rewritten by G. Masters.

One may note that the standard seismic parameters (V_P, V_S and ρ) are not directly inverted, but are used, as an intermediate stage, to link the temperature and the anisotropy (for a given mineralogy) to the surface wave data.

2.1.2 Anisotropy

The other parameter of our models is the depth-distribution of the radial anisotropy, which is constrained by the Rayleigh–Love discrepancy (Anderson 1961). Seismic anisotropy may have various origins, like lattice or shape-preferred orientation of mantle minerals (Mainprice 2007), and/or a superposition of alternating layers with highly contrasting seismic properties (Backus & Gilbert 1962). In a transversely isotropic medium, the number of independent elements of the fourth-order elastic tensor reduces to the five Love coefficients (Love 1927), $A = \rho V_{P_H}^2$, $C = \rho V_{P_V}^2$, $N = \rho V_{S_H}^2$, $L = \rho V_{S_V}^2$ and F; where ρ is the density, V_{P_H}, V_{P_V}, V_{S_H}, V_{S_V}, are the velocities of the horizontally and vertically propagating P waves, and the horizontally and vertically polarized S waves, respectively. Since the behaviour of these anisotropic *moduli*, as a function of temperature and pressure, is still unknown for most mantle minerals, we follow the procedure of Babuska & Cara (1991), Panning & Romanowicz (2006) and Khan *et al.* (2009), and define an isotropic shear wave velocity V_S by the Voigt average

$$V_S^2 = \frac{2V_{S_V}^2 + V_{S_H}^2}{3} = \frac{2L + N}{3\rho}. \tag{1}$$

The anisotropy parameters are

$$\xi = \frac{V_{S_H}^2}{V_{S_V}^2} = \frac{N}{L}, \tag{2}$$

and

$$\eta = \frac{F}{A - 2L}. \tag{3}$$

Since the Rayleigh–Love discrepancy is mostly sensitive to the shear wave anisotropy (Babuska & Cara 1991), we discard P-wave anisotropy and set $\eta = 1$. Eqs (1) and (2) are combined to compute V_{S_H} and V_{S_V} as a function of V_S and ξ,

$$V_{S_H} = V_S\sqrt{\frac{3\xi}{2+\xi}}, \tag{4}$$

$$V_{S_V} = V_S\sqrt{\frac{3}{2+\xi}}. \tag{5}$$

[1] Computational Infrastructure for Geodynamics, http://www.geodynamics.org/cig/software/mineos

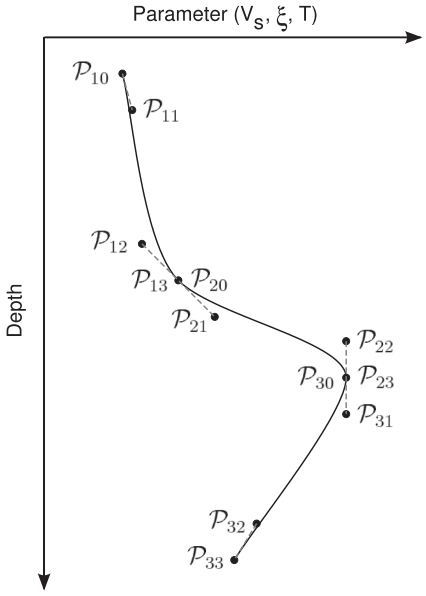

Figure 1. Schematic representation of the model parametrization using a continuous set of C^1 Bézier curves. Four control points define each polynomial and its derivative. Continuity requires that the upward and downward derivatives are identical at each point that is common to two consecutive polynomials. It works for any kind of parameter.

2.2 Model parametrization

The temperature and anisotropy distributions are described, as a function of depth, using C^1 Bézier polynomials (Bézier 1966, 1967), based on randomly chosen control (or anchor) points. This procedure offers several advantages: (1) it does not impose a regularly spaced discretization of the models or *a priori* on layer thicknesses and location of the seismic discontinuities; (2) it can be used to describe both smooth (e.g. temperature gradients) and sharp (e.g. thin thermal boundary layers) variations with a minimum number of parameters; and (3) it may be optimized during the iterative processes of the McMC algorithm by adapting the number of curves that are necessary to describe a given structure to the resolving power of the observations.

The parametrization of a temperature or anisotropy distribution is sketched in Fig. 1. The overall distribution is described by a series of N elementary cubic polynomials $B(t)$ of the form

$$B_j(t) = \sum_{i=0}^{3} b_{i,3}(t)\mathcal{P}_{ji}, \quad t \in [0,1], \quad j = [1 \ldots N], \tag{6}$$

where the $\mathcal{P}_{ji}$, $i = 0, \ldots, 3$ are the control points and where

$$b_{i,k}(t) = \binom{k}{i} t^i (1-t)^{k-i}, \quad i = 0, \ldots k \tag{7}$$

are the Bernstein basis polynomials of degree k. At the ending points $\mathcal{P}_{j0}$ and $\mathcal{P}_{j3}$, the curvature is defined by the norm of the local tangent vectors $\mathcal{P}_{j0}\mathcal{P}_{j1}$ and $\mathcal{P}_{j2}\mathcal{P}_{j3}$, respectively. The C^1 class requires that the upward and downward derivatives are identical at each point joining two consecutive polynomials.

3 McMC METHOD

Our inverse problem consists of computing the radial distributions of temperature and anisotropy from surface wave dispersion curves. This section outlines the fundamentals of the Bayesian inversion,

based on the McMC method, detailed in Mosegaard & Tarantola (1995) and Mosegaard (1998), and successfully applied to constrain the Earth's mantle structure using geophysical data (e.g. Khan *et al.* 2006; Verhoeven *et al.* 2009).

Let us denote by $\mathbf{p}$ the parameters of our model and $\mathbf{d}$ the seismological data, respectively. The data are related to the parameters through the equation, $\mathbf{d} = A(\mathbf{p})$, where the non-analytic and non-linear operator A represents the forward problem discussed in Section 2. Explicitly, the parameters are the control points of the cubic Bézier curves, namely depth (z), temperature (T) and anisotropy (ξ) at these particular points. In the Bayesian framework, a set of parameters is randomly chosen at each iteration. The corresponding radial distributions of seismic velocities and densities are used to compute Rayleigh- and Love-phase velocity dispersion curves.

The solutions of the inverse problem are described by the posterior probabilities $P(\mathbf{p}|\mathbf{d})$ that the parameters are in a configuration $\mathbf{p}$ given the data are in a configuration $\mathbf{d}$. The parameter space is sampled according to $P(\mathbf{p}|\mathbf{d})$. Bayes' theorem links the prior distribution $P(\mathbf{p})$ and the posterior distribution $P(\mathbf{p}|\mathbf{d})$,

$$P(\mathbf{p}|\mathbf{d}) = \frac{P(\mathbf{d}|\mathbf{p})P(\mathbf{p})}{\sum_{\mathbf{p}\in\mathcal{M}} P(\mathbf{d}|\mathbf{p})P(\mathbf{p})}, \tag{8}$$

where $\mathcal{M}$ denotes all the configurations in the parameter space. It exists a set of rules named priors $P(\mathbf{p})$ which defines the prior distribution, that is, the set of possible models which reduce the configuration space and represents our state of knowledge. For example, we do not know the temperature profile in the transition zone, but we can set upper and lower bounds on its expected value. Only models that belong to the prior distribution are proposed and tested against the data.

The probability distribution $P(\mathbf{d}|\mathbf{p})$ is a function of the misfit $S(\mathbf{d}, A(\mathbf{p}))$, which measures the difference between the observed data $\mathbf{d}$ and the computed synthetic data $A(\mathbf{p})$,

$$P(\mathbf{d}|\mathbf{p}) \propto \exp(-S(\mathbf{d}, A(\mathbf{p}))). \tag{9}$$

The misfit function takes into account the overall variance that we assign to the data. We assume independent and normally distributed data errors. The misfit function is evaluated with an L_2 norm:

$$S(\mathbf{d}, A(\mathbf{p})) = \sum_i \frac{(d_i - A(p)_i)^2}{2\sigma_i^2}, \tag{10}$$

where d_i and $A(p)_i$ represent the i-th component of the observed data vector $\mathbf{d}$ and of the synthetic data, respectively, and σ_i is the standard deviation associated with the i-th observation. The data variance directly determines the form of the posterior probability distribution and hence the posterior samples generated from it. In eq. (10), the sum runs over all considered surface wave modes and frequencies.

Bayes' formula provides an estimate of the marginal probability $P(p_i = x|\mathbf{d})$, which is obtained by summing all the probabilities of the $\mathbf{p}$ configurations where the i-th parameter takes the value x:

$$P(p_i = x|\mathbf{d}) = \frac{\sum_{\mathbf{p}\in\mathcal{M}_x^i} P(\mathbf{d}|\mathbf{p})P(\mathbf{p})}{\sum_{\mathbf{p}\in\mathcal{M}} P(\mathbf{d}|\mathbf{p})P(\mathbf{p})}, \tag{11}$$

where $\mathcal{M}_x^i$ represents all the configurations whose i-th component is x. The evaluation of the marginal probability is impossible to calculate in practice, because the sum in the denominator runs over the huge number of configurations. This number is equal to $\prod_{i=1}^{m} |\mathcal{M}_i|$, where $|\mathcal{M}_i|$ is the number of values that the parameter i can take. We therefore choose an McMC method, which samples the parameter space through a random walk, according to the probability $P(\mathbf{p}|\mathbf{d})$. The method is derived from the Metropolis–Hastings Algorithm (e.g. Metropolis *et al.* 1953; Hastings 1970), which generates samples according to the unknown posterior distribution. This step is done using a randomized decision rule which accepts or rejects the proposed models according to their fit to the data and the prior. After an initial exploration, the misfit function guides the research to choose another set of parameters, and the whole procedure is iterated.

Because each new model is chosen to be in a specific neighbourhood of the previous model, each new step depends only on the previous step. Let us consider the Markov chain associated with the i-th parameter with value x at the iteration t. If the prior is considered to be uniform in a given interval, the marginal probability $P(p_i = x|\mathbf{d})$ is directly equal to

$$P(p_i = x|\mathbf{d}) = \frac{\sum_{\mathbf{p}\in\mathcal{M}_x} P(\mathbf{d}|\mathbf{p})}{\sum_{\mathbf{p}\in\mathcal{M}} P(\mathbf{d}|\mathbf{p})}. \tag{12}$$

In this case, given a current model whose parameters are p_i^t and a proposed model p_i^{t+1}, a decision has to be made whether or not the new model is accepted or rejected.

(1) If $S(\mathbf{d}, P(p_i^{t+1})) \leq S(\mathbf{d}, P(p_i^t))$, the new model is accepted and added to the set of samples of the posterior distribution because it improves the fit to the data.

(2) If $S(\mathbf{d}, P(p_i^{t+1})) > S(\mathbf{d}, P(p_i^t))$, the new model is degrading the misfit function but is not necessarily rejected. The model is accepted when the ratio $P(p_i^{t+1}|\mathbf{d})/P(p_i^t|\mathbf{d})$ is larger than a random number taken from a uniform distribution between 0 and 1. Otherwise, the model is rejected.

One of the advantages of the Metropolis–Hastings algorithm is to prevent the random walk to be trapped into a local minimum. This feature can be of first importance when dealing with a non-linear problem. Another advantage is that only probability ratios are handled. Explicitly, the denominator of the eq. (8) does not need to be computed.

If the amount of iterations can be large enough, the samples provide a good approximation of the posterior distribution for the model parameters, that is, $P(\mathbf{p}|\mathbf{d})$. The posterior probability of the temperature and anisotropy at a given depth can be visualized by plotting the distribution of the selected values for the whole set of solutions.

4 PRACTICAL IMPLEMENTATION

This section details the technical aspects of the developed algorithm and its implementation to obtain the inverted synthetic and real data results shown in Sections 5 and 6.

4.1 Boundary conditions

When studying the Earth's mantle, it is well known that Rayleigh and Love surface wave data are both sensitive to the crustal structure (e.g. Montagner & Jobert 1988; Bozdag & Trampert 2007). Improper corrections may lead to the spurious mapping of crustal seismic signatures into the underlying mantle (Boschi & Ekström 2002; Ferreira *et al.* 2010; Panning *et al.* 2010). Some studies (Shapiro &

Ritzwoller 2002; Visser *et al.* 2008b; Khan *et al.* 2009; Shen *et al.* 2013a,b) attempt to circumvent this difficulty by allowing slight velocity and Moho depth variations with respect to an *a priori* crustal model.

Our region of interest is the transition zone (i.e. between 350 and 1000 km depth), and an original strategy is followed for the crust and the uppermost mantle. Due to the heterogeneous mineralogical content of the 0–350 km depth region and the inadequacy of equation of state applied at lithospheric P–T conditions, only V_S and ξ are jointly inverted above 350 km, as many tomographic studies do. This means that the stage which consists in using temperature and Birch–Murnaghan EOS and anisotropy to secondly interpret seismic models, is not implemented in this depth range. The reference model of the crust and the uppermost mantle only constrains the parameter range and its choice does not constrain the inverse scheme such as in the inversions based on perturbation theory. To assure a scale invariant parametrization (Tarantola 2005), we adopt a $\log(V_S/V_{S_0})$ parametrization, between 0 and 350 km depth. The reference velocity, V_{S_0}, can take any arbitrary value. We use scaling relations based on the experimental study of Isaak (1992) to compute V_P and ρ profiles.

From 1500 km depth down to the centre of the Earth, the seismic values are those of the PREM, but the deepest point in the inverted region is set to 1000 km. This defines a 500-km-thick region within which the continuity between the seismic profile in the transition zone and in the PREM below is ensured. This means the temperature at 1500 km depth has to be set but various numerical experiments show that this value does not influence the results in the region of interest (350–1000 km).

4.2 Prior information

The inverted parameters are the vectors corresponding to the Bézier points for shear velocity ($\mathbf{V_S}$), temperature ($\mathbf{T}$), and radial anisotropy ($\boldsymbol{\xi}$), and the depths at which these Bézier points are located ($\mathbf{z_{V_S}}$, $\mathbf{z_T}$ and $\mathbf{z_\xi}$). The vector of all inverted parameters can be written as $\mathbf{p} = [\mathbf{V_S}, \mathbf{T}, \boldsymbol{\xi}, \mathbf{z_{V_S}}, \mathbf{z_T}, \mathbf{z_\xi}]$. Various amounts of control points, which construct temperature Bézier curves (see Section 2.2), are used to allow a large sampling of the parameter space. Considering the aforementioned boundary conditions, the model vector $\mathbf{p}$ contains between 27 and 35 parameters, depending on the number of Bézier points for temperature (Table 1).

The Bayesian formulation enables to account for *a priori* knowledge, provided that this information can be expressed as a probability distribution $P(\mathbf{p})$. One feature of the Bayesian inversion is the possible tuning of the prior sampling. Here we choose minimal prior information, which consists of uniform probability distributions in wide realistic parameter spaces. Following this statement, the priors on the parameters are uniformly distributed over wide domains (Fig. 2).

For the temperature, where priors need to be set only between 350 and 1000 km depth, the low-temperature bound corresponds to the thermal state in a subducting lithosphere (Tagawa *et al.* 2007;

Table 1. Synthesis of the parameters constraining the localization of Bézier points.

	T	ξ	V_S
Number of Bézier points	6–10	7	5
Depth interval (km)	350–1000	0–1000	0–350
Minimal distance between two consecutive points (km)	80–10	40	20
Norm of the local tangent vector (km)	40–5	20	10
Maximal gradient between two consecutive points	5 K km^{-1}		

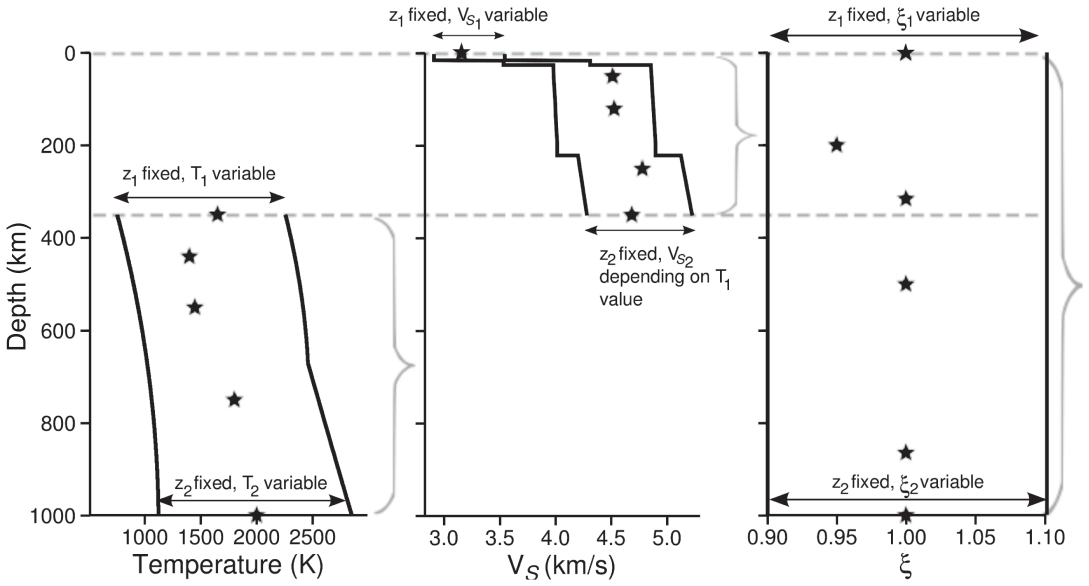

Figure 2. Illustration of *a priori* setting on the Bézier points used in both synthetic and real data inversions for temperature (T), shear wave velocity (V_S) and anisotropy (ξ). Black stars represent the Bézier points. Dotted grey lines define 0 and 350 km depths. Black lines are the bounds of the model space. Embraces indicate the depth ranges where the points are randomly sampled. The depth ranges for T, V_S and ξ profiles are 350–1000, 0–350, and 0–1000 km depth, respectively. The shear wave velocity V_{S_2} at 350 km depth is computed according to the temperature T_1.

Ganguly *et al.* 2009), while the high-temperature bound is defined by the dry solidus of a peridotite (Hirshmann 2000; Hirshmann *et al.* 2009).

For the anisotropy, inferred between 0 and 1000 km depth, the domain bounds are set to 0.9 and 1.1. These bounds are chosen to bracket the range of values obtained in recent surface wave studies (Panning & Romanowicz 2006; Kustowski *et al.* 2008; Visser *et al.* 2008b; Khan *et al.* 2009, 2011). Even when radial anisotropy might decrease with increasing depth, we do not shrink the domain range in the deepest part of the model space.

Between the surface and 350 km depth, the seismic structure is randomly sampled within the bounds of +10 per cent and −10 per cent of the PREM (Dziewonski & Anderson 1981). The seismic value at 350 km depth is not randomly sampled, but is computed from the temperature value (since it is also the shallowest point of the transition zone), owing to the thermodynamical law. To prevent from an artificial discontinuity between the uppermost mantle and the transition zone, the point located just above 350 km depth is sampled in such a way that the V_S gradient is between $\pm 2.5 \times 10^{-3}\ \mathrm{s}^{-1}$.

The Bézier points are randomly located in depth, independently for z_{V_S}, z_T and z_ξ values within the prior domains.

Figs 3(a)–(c) represent the uniform prior distributions on the Bézier control points (see Table 1 for details), and Figs 3(d)–(f) enhance the resulting prior probability distributions on the temperature, and anisotropy and uppermost V_S models, respectively, following eqs (6) and (7). Note that the continuity of the Bézier curves modifies the former uniform sampling of the parameter space using discrete control points.

To be consistent with the data resolution power, minimal distances are set between two consecutively randomly sampled Bézier control points. This leads to unsampled regions, near the depth bounds (red lines), visible in Figs 3(a)–(c).

4.3 The McMC algorithm within a three-step exploration

In the McMC algorithms, new models are proposed by randomly perturbing the previously accepted model. Here, the sampling of the parameter space is performed using a continuous proposal function and the stepsizes of the exploration depend on each parameter (V_S, T, ξ, z_{V_S}, z_T and z_ξ). For instance, defining the t-th and the (t + 1)-th value of a parameter p, as p_i^t and p_i^{t+1}, respectively, then the subsequent step may be defined as $p_i^{t+1} = p_i^t + w_i^t$, where w_i^t is the t-th stepsize, randomly sampled from a normal distribution with zero mean. A Gaussian probability density distribution, centred at p_i^t,

$$q\left(p_i^{t+1}\middle|p_i^t\right) = \frac{1}{\theta_i\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{w_i^t}{\theta_i}\right)^2\right], \tag{13}$$

is classically used to randomly sample the p_i^{t+1}, where θ_i is the standard deviation of the Gaussian distribution. If p_i^{t+1} is out of the prior bounds, then the random walk reflects on the edge to respect the equirepartition of the proposal distribution. A proper choice for the θ_i value mostly affects the efficiency of the exploration and therefore the computing time, but it does not influence significantly the posterior probability distribution (MacKay 2003). For normal distributions, an optimal jumping rule produces an acceptance rate of about 23 per cent for multidimensional problems (Gelman *et al.* 2003).

To perform a comprehensive and efficient sampling of the parameter space within a reasonable computation time, the parallel predisposition of the algorithm (Rosenthal 2000) enables us to propose an unusual split of the inversion scheme into three steps. Indeed, each chain can be conveniently placed on an independent processor of a parallel computer system. The two first steps look for a family of the best-misfit configurations of the parameters, and the statistics are performed during the third step, named 'stationary period'. The overall inversion scheme takes about 2 days on a parallel computer system. During the first step, the standard deviations (θ_i) are chosen to be large enough to explore a large part of the configuration space. A great number of chains (1600 in the presented inversions) are placed on a parallel computer system, and hence each chain runs on an independent processor. The chains start from a different state and follow different paths in the parameter space. At each iteration, the synthetic data computed from the given configuration is tested against the real data by computing the misfit (eq. 10). First, the previously described procedure (see Section 3) is used to simultaneously explore all models providing that, at each iteration, a whole set of randomly sampled trial models is tested, up to a uniform sampling of the parameter space. As stated previously, various amounts of temperature control points are used to not *a priori* fix the sampling of the parameter spaces. This step is stopped after 6×10^3 iterations and the best-misfit configuration is then determined for each independent chain.

One of the major issues when using McMC exploration is to know which iteration rank is to estimate the number of iterations needed to achieve statistical convergence of the procedure. Practically, it can be however disabling, with respect to the computation time, to stop the iterations at any time. For this reason, the second step restarts with the parameter configuration which gives the best misfit during the first step.

Only the 20 per cent best-misfit configurations of all the chains are selected for the second step to refine the results. The second step is done, again using the same McMC procedure during 6×10^5 iterations, using thinner Gaussian distributions for the parameter sampling but, this time, modifying only one couple of parameters at each iteration. The strategy of perturbing only one couple at a time ensures to preserve most of the characteristics of the current model, which may have resulted in a good data fit. This induces an increase of the acceptance rate (∼20–25 per cent) while keeping a large-scale exploration of the models. The perturbed model for the next iteration, (V_S, z_{V_S}), (T, z_T) or (ξ, z_ξ), is randomly sampled. McMC methods estimate the posterior density distribution after a 'burn-in' period, which is necessary to loose the memory of the initial configuration (starting model). The second step of our method can be considered as this 'burn-in' stage. Here the originality is to restart the Markov chain after the 'burn-in' period using the best-misfit configuration, instead of the last one.

The 'stationary period' starts after the computation of the best-misfit configurations during the second step. Only 5 per cent of these configurations, with respect to the 1600 chains considered for the first step, are engaged in the third step. Each chain runs for 6×10^5 iterations in total. The amount of Bézier control points is the same as the one randomly chosen at the beginning of the second step, and since many Markov chains are started in parallel with various amount of Bézier points, the amount of parameters for temperature is also a parameter itself, which gives some common features with the reversible jump approach (Green 1995; Green & Mira 2001; Bodin & Sambridge 2009; Bodin *et al.* 2012). Only the 4.8 millions of models sampled during the 'stationary period' are used to compute the probability density functions (pdfs). This procedure should guarantee that the samples are less correlated as well as leading to a better coverage of the probability distribution. Another advantage

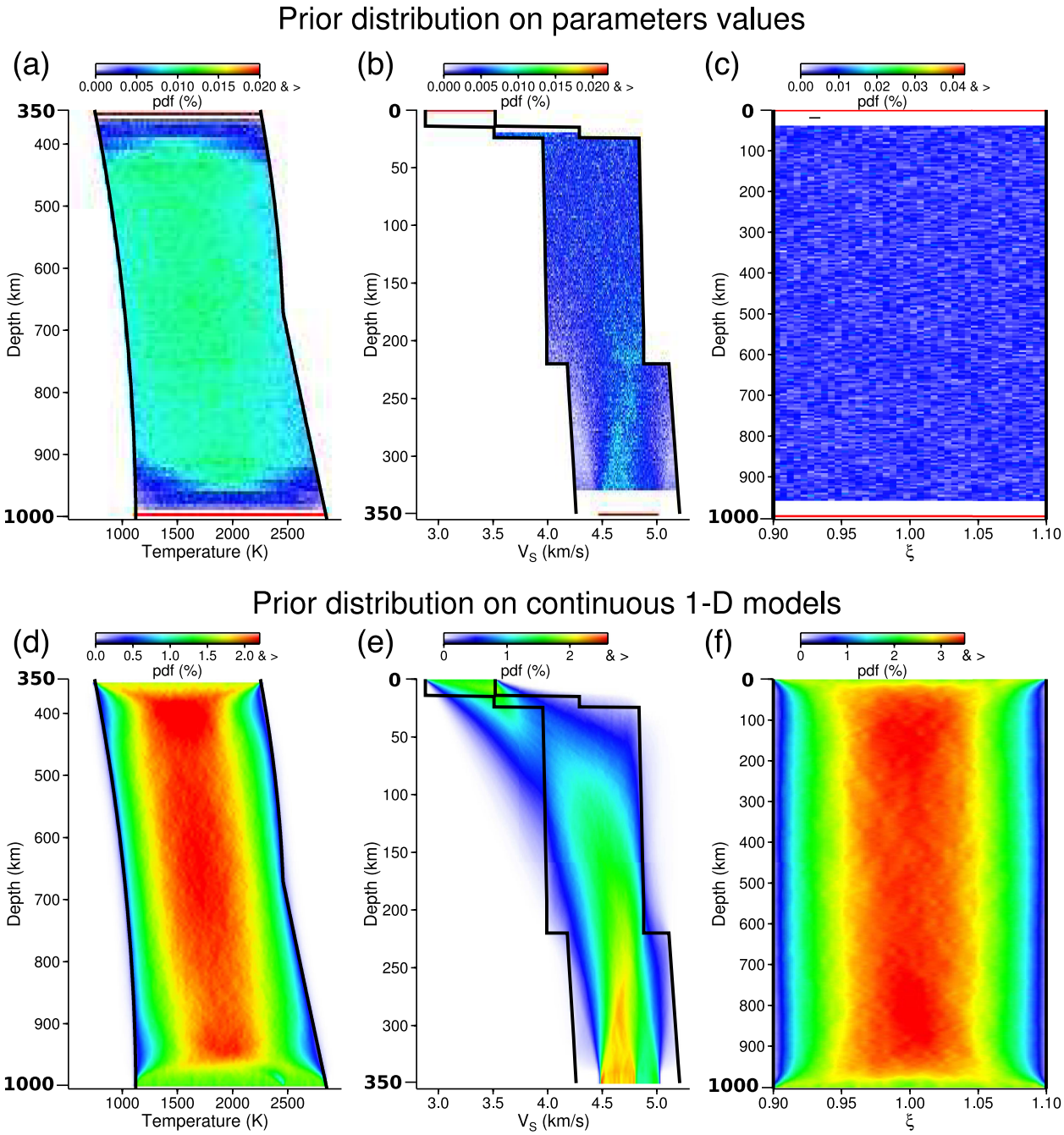

Figure 3. Marginal probability density functions (pdf) for temperature (left), shear wave velocity (middle) and anisotropy (right), considering that all the sampled models which are in good agreement with *a priori* information are accepted. The results are shown for the discrete values of the parameters (top row) and for the 1-D models (bottom row). Red and blue colours show high and low probabilities, respectively. Black lines represent the minimum and maximum parameter values allowed. The pdf values for the parameters are computed by counting the number of Bézier points in each cases. The sizes of the cases for T, V_S and ξ are 25 K $\times$ 5 km, 0.01 km s^{-1} $\times$ 2 km and 0.005 $\times$ 5 km, respectively. The pdf values for the 1-D models are computed by counting the number of curves in each parameter interval at each depth. For a given depth, the sum of the pdf over all the parameter intervals is equal to 100 per cent.

of our method is that, since the chains are preconditioned, thanks to the two first steps, the beginning of the 'stationary period' is well defined and does not need any statistical assessment.

5 INVERSION OF SYNTHETIC DATA

In this section we show how we test our method, using synthetic data. Tests on synthetic data are useful to evaluate the robustness of the method and to highlight some features which could be found on real data, considering real uncertainties on dispersion curves. We also discuss the question of how to interpret the results by the analysis of the posterior distributions. The results demonstrate the potential of the inversion method to detect and decorrelate first-order variations of temperature in the transition zone, shear wave anisotropy between the surface and 1000 km depth, and shear wave velocity in the crust and the upper mantle.

5.1 Input models

Synthetic data are computed for known temperature and anisotropy models. A pyrolitic composition is considered because it is the most widely accepted model of mantle composition, and generally compatible with a large range of petrologic, geochemical and seismic observations (Ringwood 1975). The iron number (x_{Fe} = Fe/(Fe+Mg), where Fe and Mg indicate the bulk molar abundances) is assumed to be constant throughout the whole transition zone and is set to 0.11.

To illustrate the different features of the algorithm, we choose two synthetic models represented in Figs 4 and 5. The 1-D profiles are plotted with black lines and the corresponding Bézier points with black stars. We choose temperature models far from an adiabatic profile (dashed lines in Figs 4h and 5h), to demonstrate the ability of the algorithm to detect strong temperature variations. The model 1 is created in such a way that the perturbation on phase velocities produced by temperature variations compared to an adiabatic profile is two times larger than the perturbation caused by anisotropy. Second, we investigate with model 2 the question of whether or not the anisotropy in the transition zone can be detected by putting a shear wave anisotropy anomaly at 500 km depth.

The number of parameters used and the constraints for the Bézier curves computation are shown in Table 1. To fully investigate the robustness of the method, a relative freedom is given for the placing of temperature parameters, setting the maximal gradient between two consecutive points to 5 K km^{-1}. We sample 4.8 millions of models during the 'stationary period'. To eliminate dependent samples in the ensemble solution because only one couple of parameters is modified at each iteration, every 40th visited model is selected for analysis.

Synthetic data are Love and Rayleigh surface wave dispersion curves of the fundamental mode and the first three overtones. They are computed from the input models with the forward method described in Section 2. We use in average 35 periods per mode, covering the 45–273 s period range. Realistic uncertainties taken from Beucler *et al.* (2003) are used and they vary between 0.04 and 2.4 per cent of the synthetic phase velocity values.

5.2 Characterization of posterior results

Since the posterior pdf is thought as the solution to our inverse problem, our main concern is now the analysis of the distributions. Due to the interpolation from the discrete Bézier points to the continuous 1-D profiles, uniform distributions of Bézier points do not lead to uniform distributions on 1-D profiles (Fig. 3). Consequently, results must be taken with care and other representations, in addition to colour density plots, are necessary to fully investigate the results. There is no unique way on visualizing the system under study. Thus, the results presented here reflect the particular parametrization chosen.

Before discussing the results shown in Figs 4 and 5, the four different sorts of representations used to analyse the distributions are detailed.

(1) The distribution of the Bézier points and its pdf (Figs 4a, e, i, for instance). This figure reveals the regions of the model space where the parameters are spread or narrowed. The depth of the Bézier points is not *a priori* fixed, so this representation is useful to investigate if the points are concentrated around some specific depths. The pdf gives an additional information about the most sampled regions. The drawback is that we cannot see which points are linked together to form a 1-D profile.

(2) The distribution of the Bézier points and the misfit values (Figs 4b, f, j). The information provided by the misfit is usually disregarded in Bayesian studies, as the theory relies on probabilities based on the repetition (or not) of certain values of the parameters. However, the misfit is the only quantitative expression indicating if the tested model is closed to the expected one. This figure allows to be sure that the algorithm was not trapped in a local minimum. Theoretically, the regions of low misfit values should match with regions of high probabilities.

(3) Some models randomly taken in the ensemble models and the four best-misfit models (Figs 4c, g, k). Mosegaard & Tarantola (1995) encourage this kind of representation. Here we consider a very small subset of models (20 models) and it cannot be used to infer statistical properties, but it is useful to visualize the diversity of the sampled models and to detect the depths which are the best resolved. In general, the 1-D models show different features, but all give a good fit to the data. For the synthetic tests, the profiles corresponding to the four best misfits are in general very close to expected model.

(4) The pdf of the Bézier curves (Figs 4d, h, l). This representation is the most widely employed. Using continuous Bézier curves, 1-D marginals of T, ξ and V_S can be computed at each depth. The juxtaposition of the histograms gives a 2-D density probability map. This figure provides an overview of the most frequently sampled paths, but has a tendency to smooth the results.

5.3 Results

The results of the probabilistic inversion of model 1 and 2 are shown in Figs 4 and 5, respectively. Labels are placed on figures to make easier the visualization of some particular features discussed in the following.

5.3.1 Temperature

Concerning temperature, both Bézier curves distributions contain the input models. Bézier points are scattered around the profiles, with high density values of points in the vicinity of the profiles. The misfit values are low for the Bézier points located along the profiles to retrieved. The highest density values coincide with temperature extrema (labels 1). The pdfs are sharper between 350 and 600 km depth than in the rest of the transition zone. The marginal probabilities of model 2 at 410 and 660 km depth are compared in Fig. 6. At 410 km depth, the temperature to retrieve is, without ambiguity, the most frequently sampled, while it is not so visible at 660 km depth. In Figs 4(g) and 5(g), the best-misfit profiles are all relatively close to the true models. The profiles randomly taken in the ensemble solution are oscillating around the input models between 600 and 800 km depth (labels 2) and around 900 km depth (labels 3). At these depths, the Bézier points are the most scattered. The distributions of Bézier points are narrower just above 1000 km depth due to the *a priori* PREM connection condition at 1500 km.

We performed an inversion with the same models but considering that the Earth is only composed of pure olivine, without phase transition, and we found a narrower distribution in the middle of transition zone. In this case, the pdf takes nearly the same value down to 800 km depth, suggesting that the relative broader distribution just below 600 km depth for models 1 and 2 are due to mineralogical transformations in both olivine and pyroxenes–garnet subsystems.

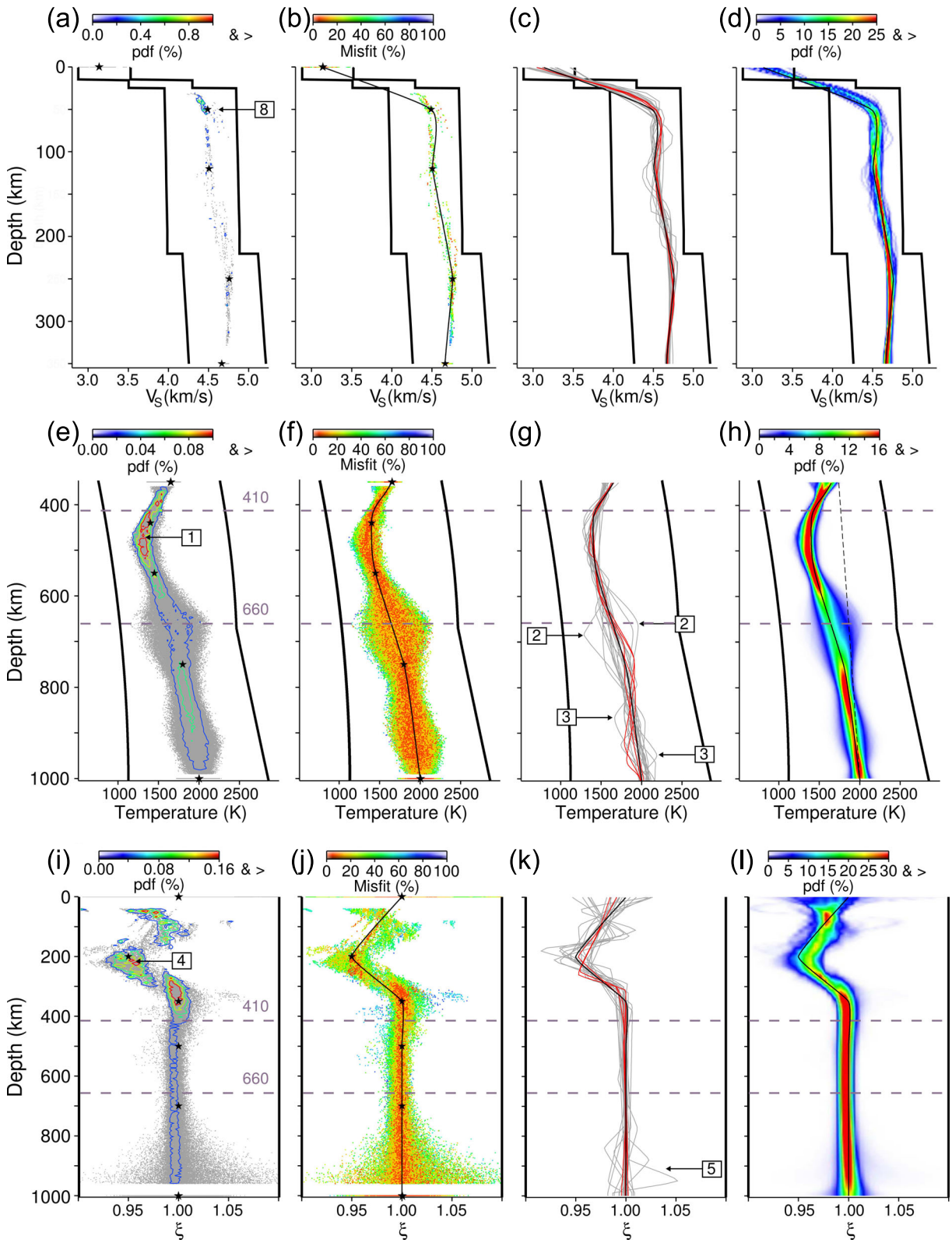

Figure 4. Results of the probabilistic inversion for the synthetic model 1 for V_S (top), T (middle) and ξ (bottom). Black lines and black stars show the models to retrieve and the corresponding Bézier points. Thick black lines are the prior bounds. (a), (e) and (i) show the Bézier points accepted during the McMC exploration. Colour lines represent the corresponding pdf. (b), (f) and (j) display the Bézier points and the misfit values. Blue and red colours are high and low misfit, respectively. Misfit values are scaled according to the highest and the lowest values encountered. (c), (g) and (k) show in grey a random subset of 20 models taken in the ensemble solution, which contains 120 000 models. Red lines are the profiles corresponding to the four best misfits. (d), (h) and (l) are colour density plots of 1-D profiles. Dashed lines in (h) show the 1600 K adiabat considering a temperature gradient of 0.4 K km^{-1}. The squared labels refer to special comment, see Section 5.3 for details.

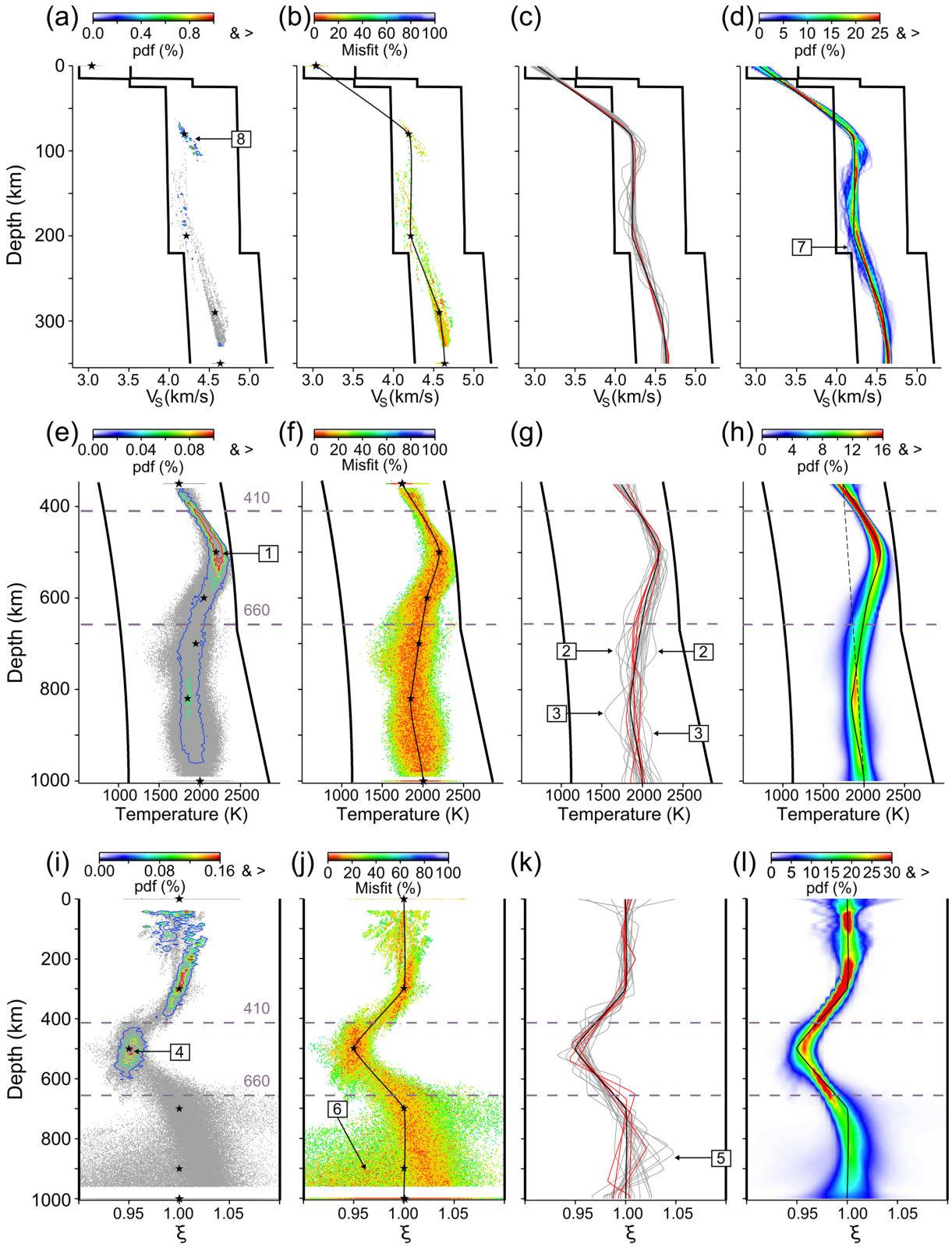

Figure 5. Same as Fig. 4 but for model 2.

In return, the oscillations below 800 km remain like in the case of a pyrolitic model, and are attributed to the decrease of surface waves sensitivity.

Fig. 7 emphasizes that models with different attributes are able to describe the same data, accounting for data uncertainties. This figure represents the pdfs of the phase velocity distributions of all the sampled models for the inversion of model 2. Fig. 7 clearly enhances that all the sampled models fit the synthetic data within their uncertainty bounds. Only the fundamental mode and overtones of Love waves are shown, but Rayleigh branches are seen to be fit

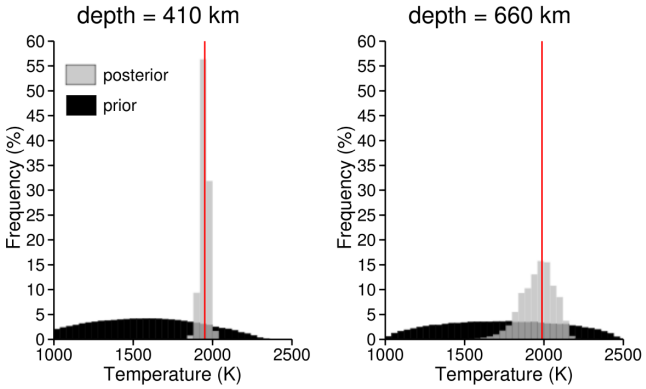

Figure 6. Prior (in black) and posterior (in grey) marginal probabilities of Bézier curves on temperature at 410 and 660 km depth for model 2. Input values are shown in red (1951 K at 410 km and 1987 K at 660 km). Intervals are 20 K wide.

within uncertainties too. Models on Figs 4(g) and 5(g) individually highlight different temperature structures of the mid-transition zone, but nonetheless are models that all produce a good fit to the synthetic data. This good agreement between synthetic and tested data reinforces the method employed here of elucidating the transition zone interior from a big number of models, rather than just adopting the 1-D profile belonging to the region of highest probability as the prevalent one (Figs 4h and 5h), and inferring the transition zone structure straight from this.

The number of Bézier points selected for each chain of the 'stationary period' is summarized in Table 2. Model 1 and model 2 are composed of 6 and 7 Bézier points, respectively. The right number of points is not the most frequently used. The percentage of sampled models is nearly equivalent for each number of Bézier points, which means that the number of points is not decisive because the parameters' depth is not *a priori* fixed. When the number of points is too high than the expected one, Bézier points are placed along the Bézier curve to retrieve and the form of the 1-D profile expected is preserved.

Table 2. Number of Bézier points selected for the 'stationary period' (per cent).

Number of Bézier points	6	7	8	9	10
Model 1 (6 points)	18.75	17.50	23.75	27.15	12.50
Model 2 (7 points)	22.50	20.00	18.75	18.75	20.00
Vanuatu–California	22.50	16.25	25.00	18.75	17.50

5.3.2 *Radial anisotropy*

Figs 4(l) and 5(l) show that the distributions of Bézier curves are in very good agreement with the expected profiles. Both anisotropy perturbations at 200 and 500 km depth, for models 1 and 2, respectively, are retrieved at the right depths (labels 4). Figs 4(j) and 5(j) show two clusters of low misfit values at these depths, which means that dispersion curves are sensitive to anisotropy variations in the upper mantle as well as in the transition zone.

The subset of 1-D models is nearly similar to the input models down to 750 km depth for model 2. Below this depth, some 1-D models are oscillating around the value of one (labels 5) and the Bézier points fill nearly the entire parameter space. Even though some points are far from the Bézier curves to retrieve, the corresponding misfit values can be relatively low (label 6). The same behaviour is observed for model 1, where the points are getting wider as a function of depth symmetrically around $\xi = 1$. According to sensitivity kernel theory, the fundamental mode and the first overtone are very sensitive to the first 200 km of the Earth. Higher modes enable to enhance the resolution down to 1000 km depth, but to a lesser extent, which explains the broader distribution of Bézier points below 700 km depth, as for temperature. Our results demonstrate that the probabilistic inversion method employed here

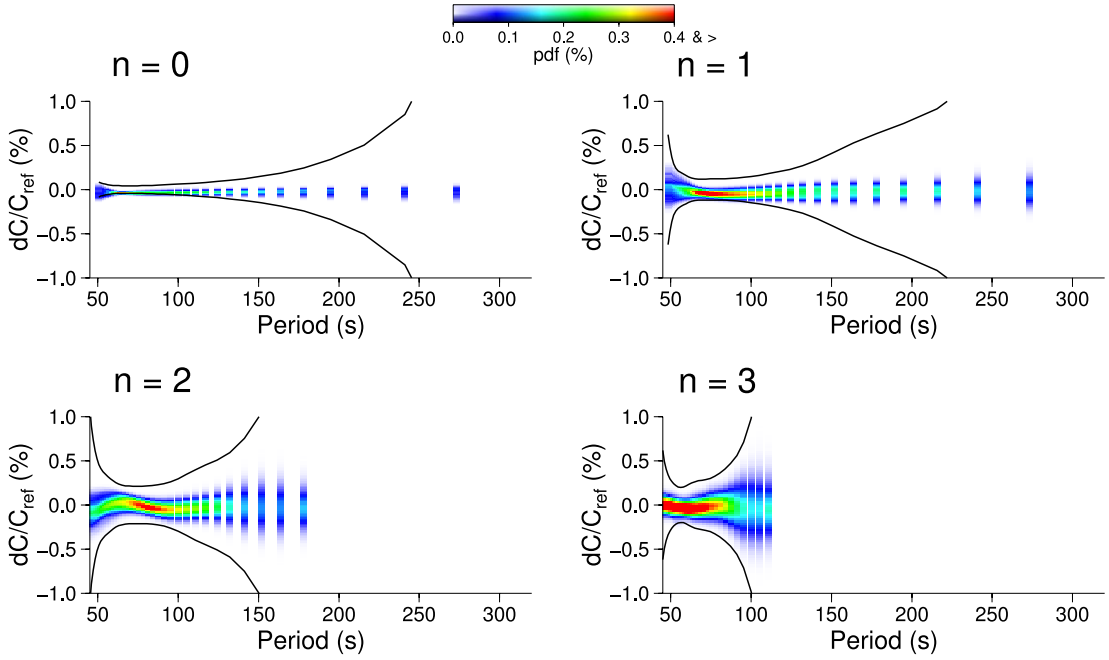

Figure 7. Posterior phase velocity distributions of Love fundamental mode and overtones of model 2. Results are shown in percentage of phase velocities (C) compared with phase velocities of model 2 (C_{ref}). Black curves are uncertainties on phase velocities. For long periods, uncertainties are large and are not displayed.

presents a clear potential to detect radial anisotropy in the transition zone, if any.

5.3.3 V_S in the uppermost mantle

Figs 4(d) and 5(d) show that V_S profiles in the uppermost mantle are well retrieved for both cases, even when the input profile is close to the prior bounds, like for model 2 (label 7). Strong changes of V_S gradient are well characterized by clusters of Bézier points. At 50 km (model 1) and 80 km depth (model2), where the slope change is the strongest, we observe a correlation between the depth of the Bézier points and V_S values (labels 8). This trade-off means unsurprisingly, that the data are fit equally well when the discontinuity is deeper and V_S is higher, and vice versa. This illustrates the limitation of the resolving power of the data. The other points are located along the Bézier curves to retrieve.

The points are more concentrated than for T and ξ results and are less numerous, which means that V_S parameters were already close to the expected values at the end of the second step of the three-step exploration (see Section 4.3), and that the algorithm did not need to change them too much during the 'stationary period'. The well-constrained V_S profiles are in good agreement with the fundamental mode sensitivity to the shallowest part of the Earth. Fig. 7 shows that fundamental mode error bounds are very thin (between 0.04 and 0.065 per cent) within the 50 to 100 s period range, which means that the misfit is significantly improved when the V_S profile is close to the true one.

6 REAL DATA

Our non-linear algorithm is used to determine the average temperature and anisotropy distributions under the Vanuatu–California path, which has been the focus of several previous studies (e.g. Cara 1979; van Heijst & Woodhouse 1997; Beucler *et al.* 2003).

6.1 Data and *a priori* conditions

The data set we analyse is made of Love and Rayleigh phase velocities up to the third order of Visser *et al.* (2008a). Phase velocity measurements were obtained using a model search approach (Visser *et al.* 2007), yielding consistent uncertainties between all the measurements. We use a total of 107 distinct phase velocities as function of period that we invert jointly for a thermal and radial anisotropy structure along a great circle. The period range investigated is 35–175 s and uncertainties are between 0.29 and 0.77 per cent of phase velocities. The bulk and shear dissipations are specified as in PREM. Pressure profile remains fixed to the PREM values, and a pyrolite composition is considered. The expected temperature gradient in the mantle is 0.4 K km^{-1} (Katsura *et al.* 2010), but we let a large freedom in the location of the Bézier points by fixing the value of $\pm$1.5 K km^{-1} to the maximal gradient between two consecutive points.

Prior boundaries are slightly different of the ones used with synthetic tests. Considering previous studies, a strong positive radial anisotropy is expected under the Pacific Plate (Ekström & Dziewonski 1998; Kustowski *et al.* 2008; Nettles & Dziewonski 2008; Khan *et al.* 2009). Consequently, we extend the upper bound of ξ to 1.3 instead of 1.1. Because the path under study is oceanic, high seismic values are expected in shallow layers and we allow a wider range of seismic velocities for the first 25 km. Setting relatively large bounds prevents from compensation effects between the different sets of parameters because several parameters cannot reach adequate values.

6.2 Results

6.2.1 Temperature distribution

Figs 8(e)–(h) show that the temperature distribution is globally close to an adiabat. As for synthetic tests, the Bézier points are spread at all depths and the profiles do not depend on the number of points used (see Table 2). The temperature is particularly well defined between 350 and 550 km depth, where the pdfs are the highest (Figs 8e and h). The distribution is more scattered in the mid-transition zone, between 550 and 750 km depth.

In Fig. 8(g), three kinds of profiles can be distinguished. The first family (referred as profiles 1 in red) is nearly adiabatic and is very close from the median of the distribution (in blue). The temperature gradient of the median profile is about 0.22 K km^{-1}, which is slightly lower than the supposed one in the mantle. When the temperature values are extrapolated to the surface, a potential temperature of about 1600 K is obtained. The second (in green) and third (in purple) families of models are characterized by an increase and a decrease of temperature, respectively, with respect to the median profile.

Although the pdf of Bézier curves appears to be smooth and unimodal between 550 and 750 km depth (Fig. 8h), the statistic results on Bézier points (Fig. 8e) enhance two distinct clusters (labels 1 and 2), also visible with misfit values (Fig. 8f). They can be assimilated to the families 2 and 3. All the results gathered together, it seems that both high and low temperatures in the mid-transition zone are able to fit the data within uncertainties, as shown in Fig. 9.

6.2.2 Radial anisotropic structure

Between the surface and 350 km depth, a strong positive ($V_{SH} > V_{SV}$) anisotropic signal nearly equal to 1.05 is revealed. It seems to be a robust feature because Bézier points are all concentrated around this value (Figs 8i and j). This result is different from the PREM but the shear wave anisotropy structure retrieved here concurs with upper-mantle results obtained from recent whole mantle seismic tomographic studies. The origin of this anomaly is not well understood, though several authors have proposed that this region of anisotropy corresponds to a flow channel at the boundary between the lithosphere and asthenosphere (Gaboret *et al.* 2003; Gung *et al.* 2003). The apparently strong anisotropy that is depicted near the surface might be due to the inability of surface waves to resolve the details of shallow crustal layers at low frequencies (Backus & Gilbert 1962; Fichtner *et al.* 2013). Between 350 and 450 km depth, the signal amplitude of anisotropy is decreasing as a continuous fashion as the transition zone approached. The distribution is centred around a value of one down to 1000 km depth. As we observed with synthetic tests, the distribution of Bézier points is broader below 700 km depth, due the decrease of surface waves sensitivity.

Several models (Fig. 8k) show slight oscillations around $\xi = 1$ at depths deeper than 400 km (labels 3 in Fig. 8k and 4 in Fig. 8j). The marginal distributions of ξ at this depth is represented for Bézier points and Bézier curves in Fig. 10. Both histograms clearly show that $\xi \sim 1$ is the most probable value. The case for a slight ($\sim$0.97 at most) negative shear wave anisotropy ($V_{SH} < V_{SV}$) below 400 km depth, as observed by Montagner & Kennett (1996), Panning & Romanowicz (2006) and Visser *et al.* (2008b), is not substantiated here and our results indicate that the transition zone is nearly isotropic.

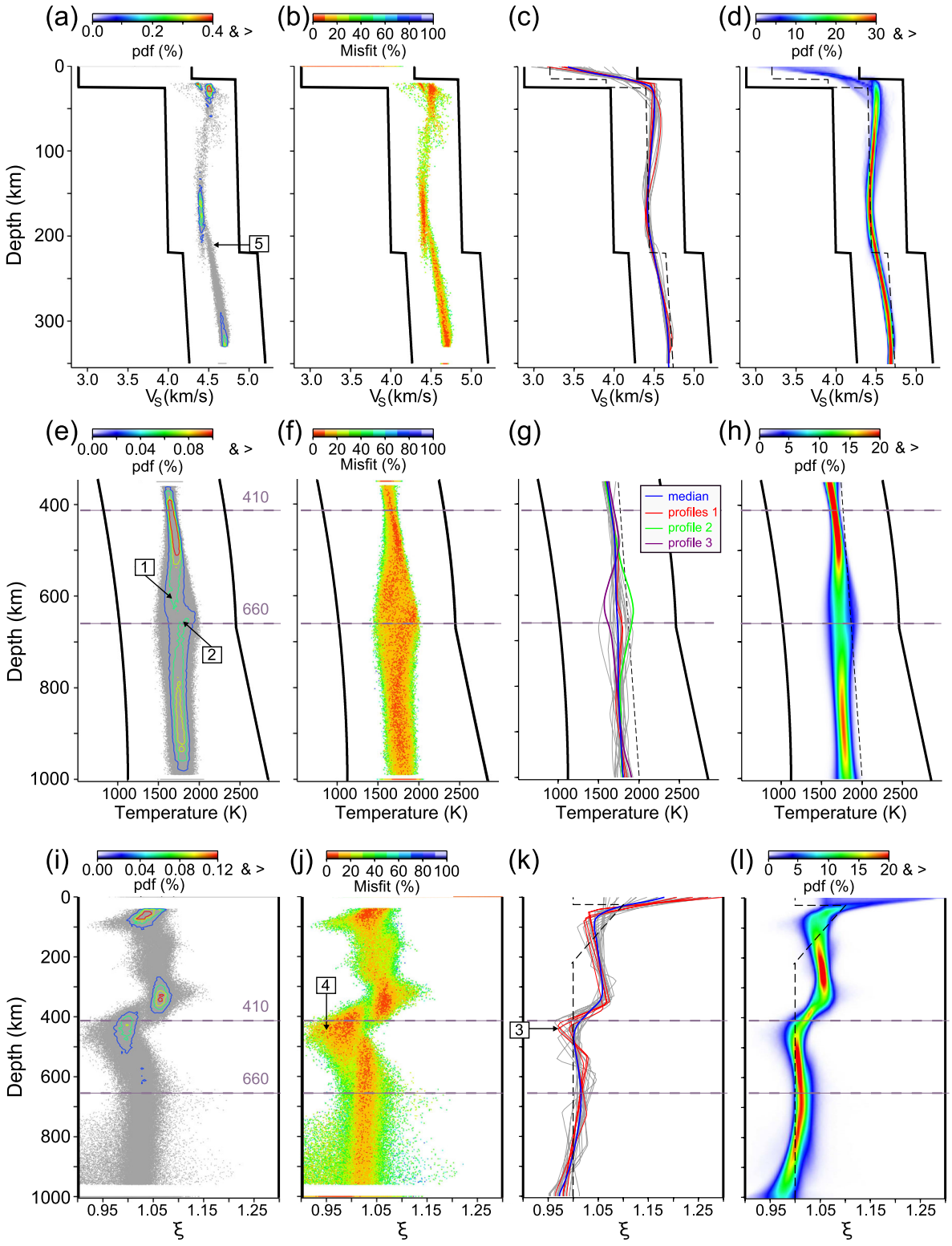

Figure 8. Same as Fig. 4 but for the Vanuatu–California data. Dashed lines in (c), (d), (k) and (l) are PREM values. Dashed lines in (g) and (h) show the 1600 K adiabat considering a temperature gradient of 0.4 K km^{-1}. Blue lines in (c), (g) and (k) are the median profiles of the distributions.

6.2.3 Seismic velocities and density

We describe as in the synthetic tests, the results of the primary parameters of the inversion scheme, that is, V_S between the surface and 350 km depth (Figs 8a–d). Moreover, the derived physical properties from the temperature distribution are discussed, in the form of *S*- and *P*-wave velocities and density distributions of the transition zone (Fig. 11), since these parameters are also determined during the whole direct problem.

V_S in the uppermost mantle. At the surface, V_S values are globally large compared to PREM and are consistent with an oceanic path. Unlike synthetic tests, all depths are sampled with Bézier points

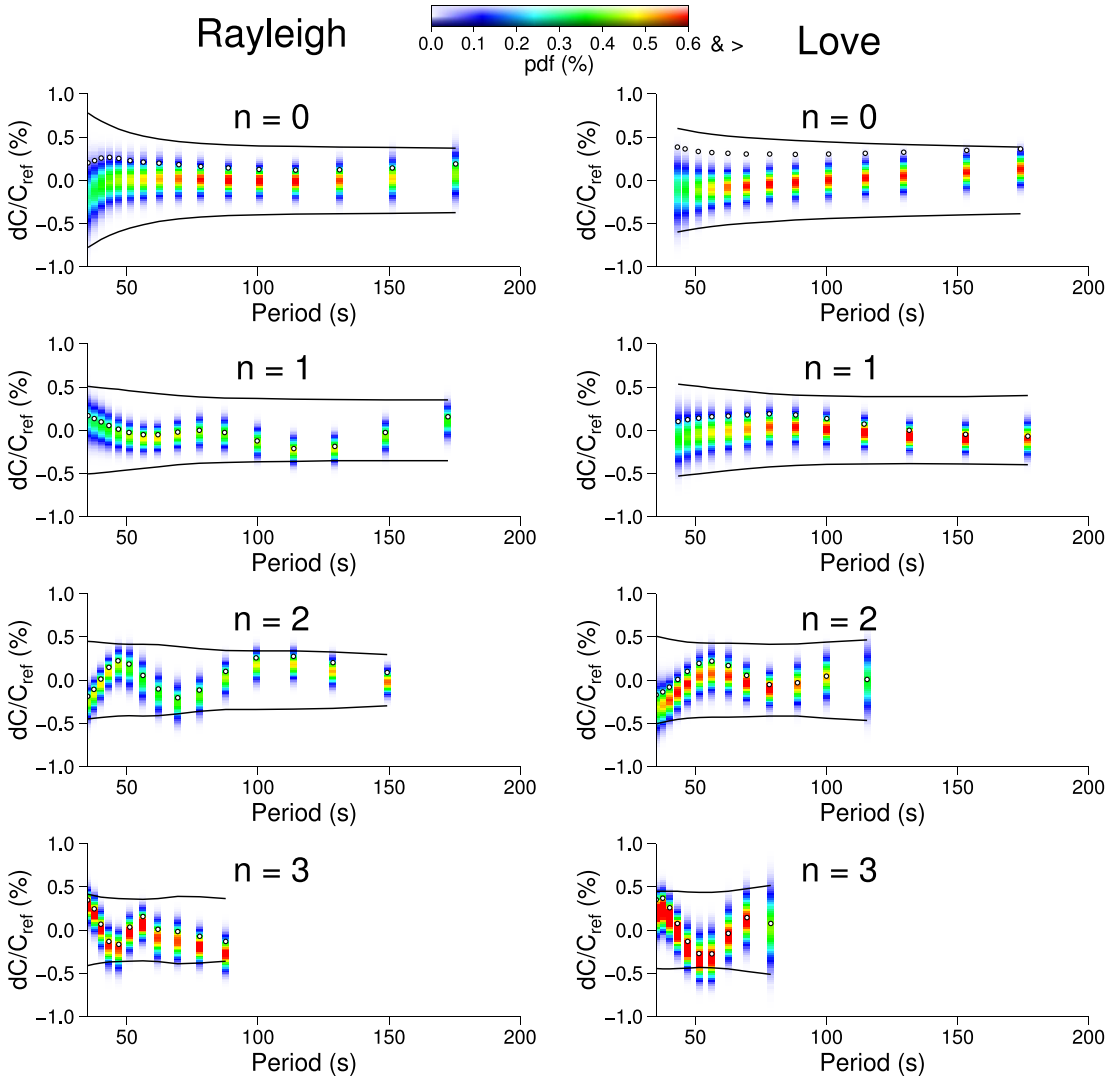

Figure 9. Same as Fig. 7 but for the Vanuatu–California data. White dots are phase velocities of the median profiles plotted in Fig. 8.

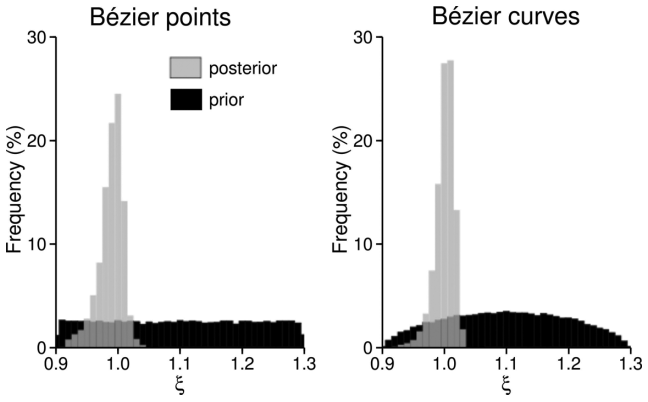

Figure 10. Marginal prior (in black) and posterior (in grey) distributions of Bézier points and Bézier curves for shear wave anisotropy at 450 km depth. It concerns the Vanuatu–California path. Intervals are 0.01 wide.

(Figs 8a and b). This is due to larger data uncertainties, which allow the algorithm to accept wider configurations of Bézier points.

Between 50 and 150 km depth, V_S is smoothly decreasing, and can be considered to be a low-velocity zone extensively described before under oceans (e.g. Gaboret *et al.* 2003; Gung *et al.* 2003; Nettles & Dziewonski 2008). A change slope that we assimilate to the Lehmann discontinuity occurs between 200 and 300 km depth. This feature is clearly visible on Bézier points, where they split into two branches (label 5). Precise agreement is not warranted here as surface wave data are not as sensitive to the exact location of discontinuities as are converted or reflected phases, for example.

V_S, V_P and ρ in the transition zone. The V_S, V_P and ρ distributions plotted in Fig. 11 are computed using the forward problem as detailed in Section 2.1.

In the upper half of the transition zone (between 400 and 520 km depth), *P*- and *S*-wave velocity distributions are found to be larger than the PREM, while in the lower part *P*- and *S*-waves velocities are slightly lower, implying an overall steeper velocity gradient in this region. The density distribution agrees with PREM between the transformations ol→wad and wad→ring. At the latter, density increases and become more than PREM.

The depth where the phase transition ol→wad occurs is similar to the one of PREM, roughly at 400 km depth. The temperature the most frequently sampled at this depth is ∼1650 K, within the range that is expected for this transition from high-pressure experiments, that is, ∼1750 ± 100 K (Ito & Takahashi 1989). A striking feature is the magnitude of the seismic jump near 400 km depth of our V_S and V_P distributions, relative to PREM. This discrepancy between

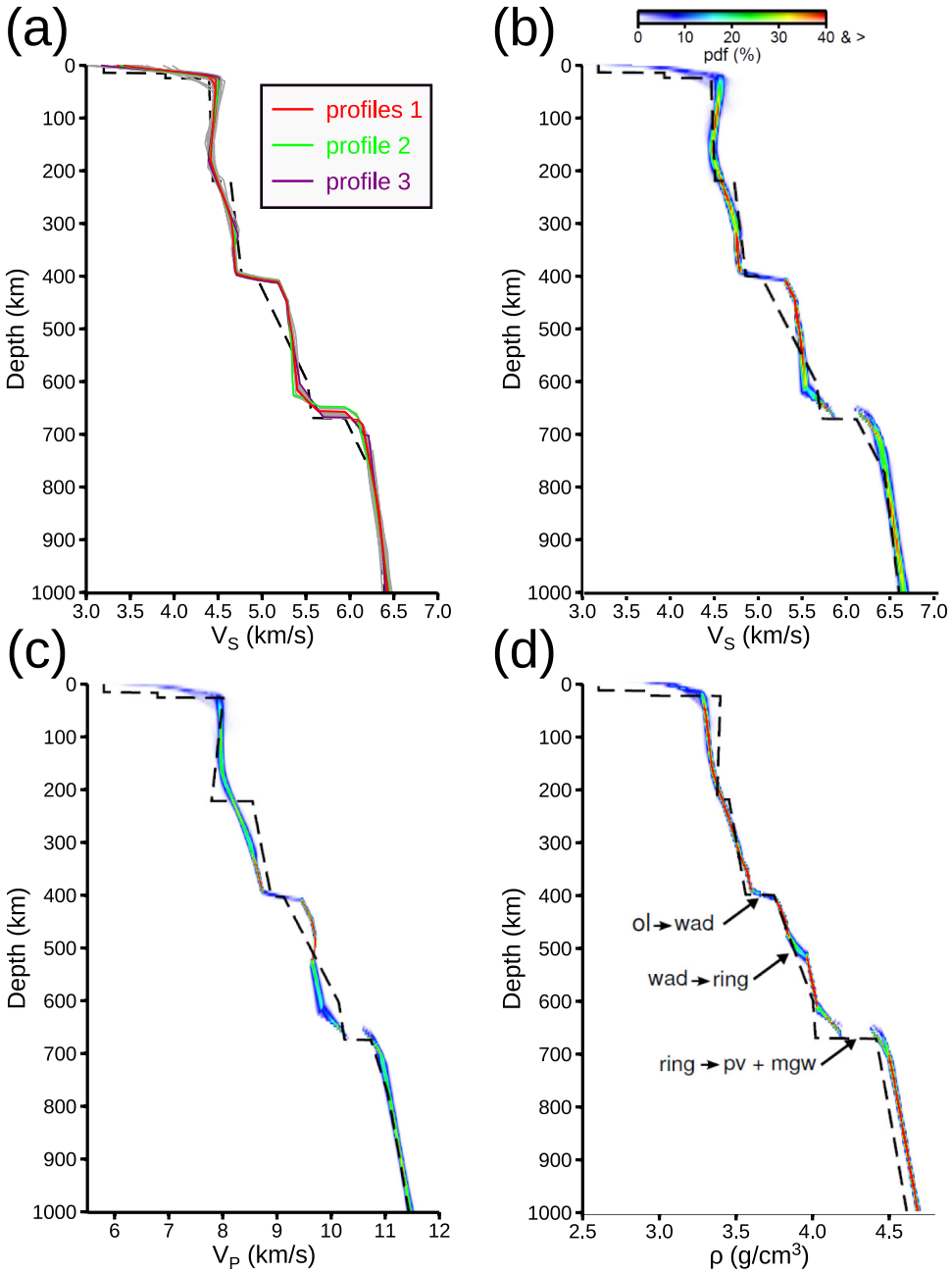

Figure 11. Probability maps of (b) V_S, (c) V_P and (d) ρ as a function of depth for the Vanuatu–California data. The abbreviations in (d) stand for ol, olivine; wad, wadsleyite; ring, ringwoodite; pv, perovskite; mgw, magnesiowustite: (a) shows in grey a random subset of 20 models in the ensemble solution and in colour the four models corresponding to the best misfits. Dashed lines are PREM values.

the results based on elastic modulii taken from laboratory experiments and global models was already noted by Stixrude (1997), Cammarano *et al.* (2003) or Katsura *et al.* (2004) for a pyrolitic composition. The amplitude of the discontinuity could be reduced considering a less olivine-rich mineralogical model (Duffy *et al.* 1995). Instead, the amplitude of the seismic discontinuities may be underestimated in the global seimic models (Shearer 2000), which may be due to finite frequency effect of seismic waves (Jackson 1998). However, as the data considered here are less sensitive to the location and size of the discontinuities, we leave it for further study to investigate this discrepancy.

The phase transition wad→ring (the 520-km discontinuity) is not present in PREM. There is no jump near 520 km depth concerning the V_S distribution, but it appears in the density distribution, where the jump occurs between 490 and 520 km depth. The V_P distribution shows a slightly low-velocity zone around 500 km depth. This behaviour is linked to the particular values chosen for the wadsleyite and ringwoodite elastic modulii and their derivatives with respect to temperature and pressure, and was already noted by Cammarano *et al.* (2005).

Between 600 and 700 km depth, the distributions are more complicated. The inverse problem is highly non-linear, and hence the posterior is far from being an unimodal Gaussian distribution. To illustrate this, we plot in Fig. 12 the marginal distributions for V_S at 655, 660 and 665 km depth. The 660 km histogram has two maxima, which means that the marginal distribution is influenced by

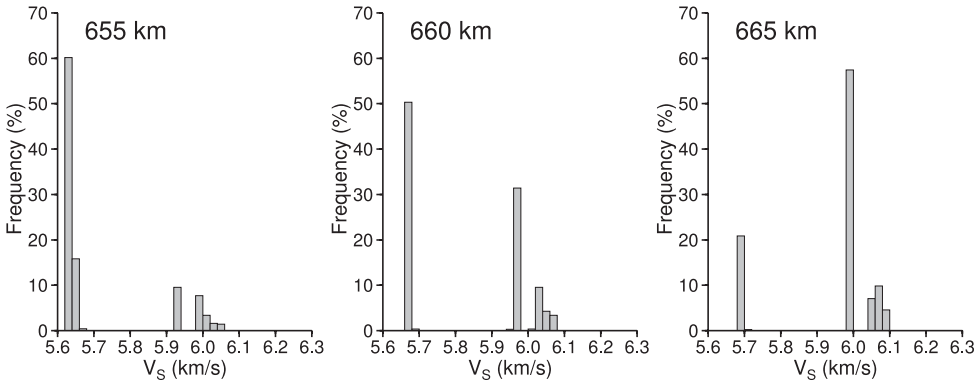

Figure 12. Marginal posterior distributions for V_S at 655, 660 and 665 km depth (i.e. around the transformation of ringwoodite into perovskite and magnesiowustite). Intervals are 0.02 km s^{-1} wide. The distribution is clearly influenced by both V_S values taken above and under the discontinuity.

velocity values above and below the discontinuity. Looking at Fig. 11(a), the mineralogical transformation of ringwoodite to perovskite and magnesiowustite occurs over a wide depth range above the PREM discontinuity. In Fig. 11(a) are plotted the V_S profiles corresponding to the three families of temperature profiles shown in Fig. 8(g). Because the ring→pv + mgw transformation is endothermic (e.g. Ito & Takahashi 1989), the coldest model (profile 3) has the deepest discontinuity, and the hottest (profile 2) has the shallowest one.

Just below 700 km depth, an additional transition to higher velocities is observed on several models in Fig. 11(a), due to other phase transformations in the pyroxenes–garnet subsystem. Deeper, *S*- and *P*-wave velocities are equal or slightly higher than PREM, but the density is higher by about 1.7 per cent.

6.3 Effect of the Clapeyron slope at 660 km depth on the temperature distribution

The distribution of the Bézier points defines three possible profiles of temperature. They are displayed in Fig. 13, along with stability range of the olivine polymorphs (in black) and Mg-perovskite and magnesiowustite (in pink). The limit between the latter domains clearly coincides with the Clapeyron slope (dP/dT) used in this study, $\Gamma_{660} = -2.8$ MPa K^{-1} (Ito & Takahashi 1989).

To illustrate the covariance between temperature and composition, we further investigate the effect of the Clapeyron slope on the temperature distribution, by setting its value to $\Gamma_{660} = 0$ MPa K^{-1}. Then, the transformation occurs instantaneously when the temperature profiles cross the limit of 24 GPa (Fig. 13). The results for V_S and ξ distributions are very similar to the ones obtained with $\Gamma_{660} = -2.8$ MPa K^{-1} (not shown here), and no trade-offs between V_S, T and ξ were obtained. The temperature results are displayed in Fig. 14. Although the distribution is similar to the one of Figs 8(e)–(h) down to 500 km depth, a $\Gamma_{660} = 0$ MPa K^{-1} implies a temperature decrease in the mid-transition zone. If the best-misfit profiles (Fig. 14c) are compared to the three families of profiles (Fig. 13), we observe that only family 3 is preserved. The results of this test enhance that the temperature values that are inverted close to seismic discontinuities depend on the priors on the composition and thermodynamical parameters.

7 CONCLUSION

In this paper, we present a new method combining thermodynamical laws with a non-linear inversion scheme to infer radial distributions

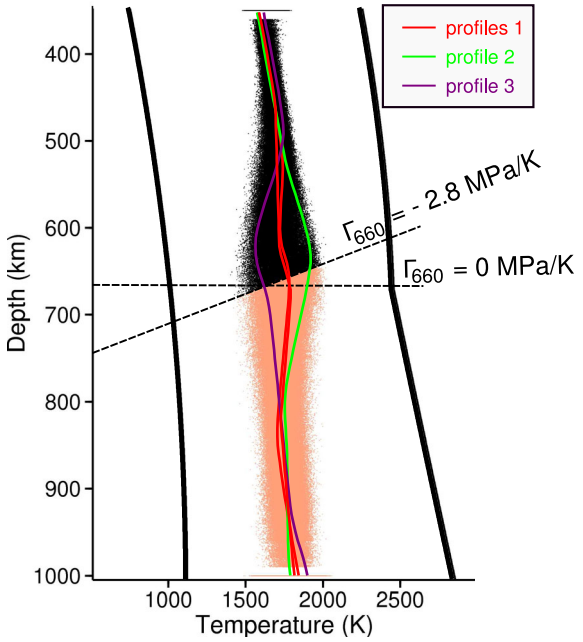

Figure 13. Distribution of Bézier points for temperature. When the thermodynamic conditions favour the olivine polymorphs, the points are plotted in black, and when perovskite and magnesiowustite are created, the points are shown in pink. Γ_{660} refers to the Clapeyron slope value.

of temperature in the transition zone, shear wave anisotropy in the mantle and shear wave velocity in the crust and in the upper mantle from seismic surface wave data. This method was used to invert fundamental and higher-order Love and Rayleigh dispersion curves (up to the third overtone). This was done with an McMC algorithm, that is, by performing a random walk in a multidimensional model space that combines prior information with misfit to the data from model parameters. Input to the algorithm were random models generated according to the prior distribution and the likelihood function. As output we assimilated random realizations of the posterior distribution. We chose to invoke as few prior constraints as possible to gauge which particular feature is most probable, given data and *a priori* information. In our approach, size and location of discontinuities in seismic properties are modelled in a physically realistic manner, as their variations depend on composition and physical conditions of the particular model being considered. The use of Bézier points

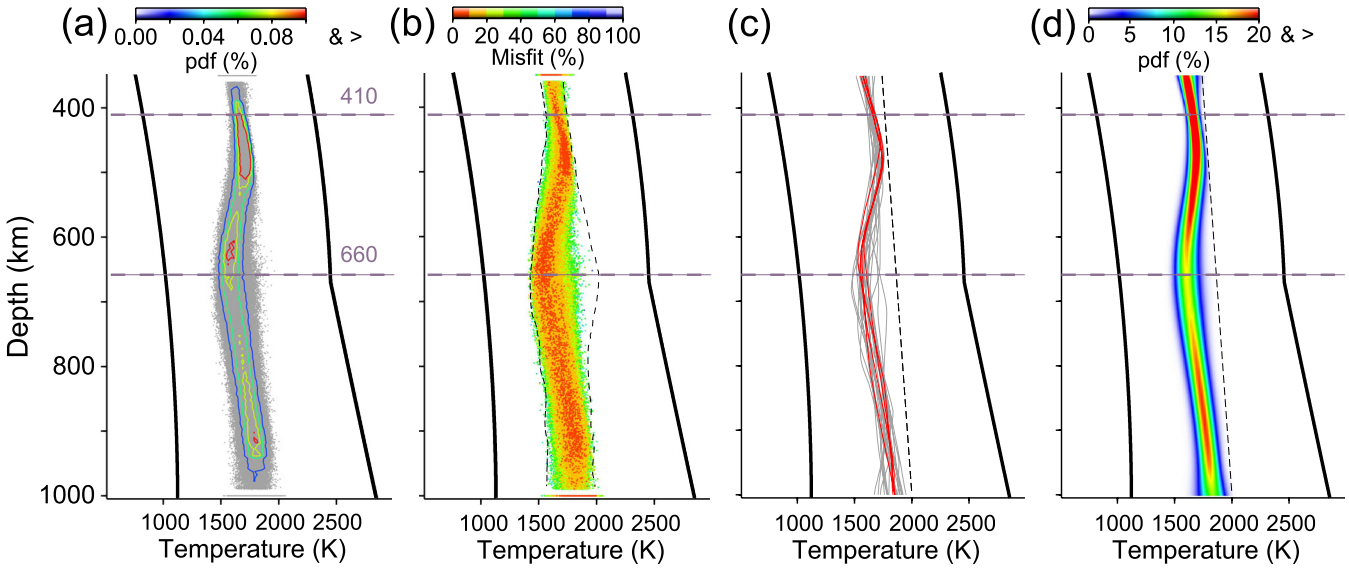

Figure 14. Same as Fig. 8 but for the Vanuatu–California data, considering a Clapeyron slope equal to zero for the ringwoodite to perovskite and magnesiowustite transformation. Dashed lines in (b) are the contours of the Bézier points distribution displayed in Figs 8(e) and (f).

as models parameters, whose depths are not *a priori* fixed, constitutes an adaptive parametrization that allows a great flexibility in the exploration of the model space.

We tested our algorithm with synthetic and real phase averaged velocity dispersion curves. We furthermore give some examples of how to investigate the sampled models. The results show that the data are sensitive to temperature and shear wave anisotropy in the transition zone. As for radial anisotropy, the picture that emerges of the Visser *et al.* (2008a)'s data is in overall accordance with other studies, that is a strong shear wave anisotropy in the upper mantle and a roughly isotropic behaviour in the transition zone. Improvements on phase velocities uncertainties are surely welcome to discriminate whether or not a significant sign change really occurs below 400 km depth. The temperature distribution is nearly adiabatic with a gradient of 0.22 K km^{-1} and a potential temperature of 1600 K, but is relatively complex between 600 and 700 km depth, which means that the thermal structure of the transition zone cannot be reduced to a unique 1-D profile. In addition, we observed that the temperature distribution is very sensitive to the choice of the Clapeyron slope of the ring$\rightarrow$pv + mgw transformation.

The range of reported Clapeyron slopes for the perovskite-forming reaction is wide: -3.0 (Ito & Takahashi 1989) to -0.2 MPa K^{-1} (Litasov *et al.* 2005). A further study would be to evaluate the effects of the Clapeyron slope on our inferred temperature distribution by considering Γ_{660} as a parameter in itself. We tested for a joint inversion of temperature and mineralogy for synthetic data, but posterior distributions of volume fractions of mantle minerals was always very close to the prior distributions. Olivine and pyroxenes exhibit a nearly identical mechanical behaviour, and cannot be distinguished on the basis of seismic data only. Given the generality of our formulation, additional independent data sets such as electrical conductivity and anelasticity properties, can easily be incorporated to separate the effects of temperature and mineralogy (e.g. Verhoeven *et al.* 2009).

McMC methods in seismology are promising, and offer a nice alternative to deterministic methods which tend to find a model which lies close to its start. The main restriction is the computational cost. In spite of this limitation, since the proposed algorithm is a direct parameter search, the forward calculations are separate routines independent of the main algorithm, and hence they can be easily replaced by alternative algorithms.

ACKNOWLEDGEMENTS

The authors are grateful for constructive reviews by Lapo Boschi and Thomas Bodin, and the Associate Editor Michael Ritzwoller. Discussions with A. Rivoldini were really fruitful. The authors like to extend their gratitude to the people who allowed them to use the two centres of intensive computing for running their algorithm. They thank the team of CCIPL (Centre de Calcul Intensif des Pays de la Loire). This work was also performed using HPC resources of CINES (Centre Informatique National de l'Enseignement Supérieur) under the allocation 2013047011 made by GENCI (Grand Equipement National de Calcul Intensif). Except Figs 1 and 2, all figures are produced by the GMT graphics software (Wessel & Smith 1998). This research has been funded by the ANR-TAM (BLANC08-1_323200).

REFERENCES

Anderson, D.L., 1961. Elastic wave propagation in layered anisotropic media, *J. Geophys. Res.,* **66,** 2953–2963.

Babuska, V. & Cara, M., 1991. *Seismic Anisotropy in the Earth,* Kluwer Academic Publishers, p. 271.

Backus, G.E. & Gilbert, J.F., 1962. Long-wave elastic anisotropy produced by horizontal layering, *J. geophys. Res.,* **67,** 4427–4440.

Beghein, C., Resovsky, J. & van der Hilst, R.D., 2008. The signal of mantle anisotropy in the coupling of normal modes, *Geophys. J. Int.,* **175,** 1209–1234.

Beucler, É., Stutzmann, É. & Montagner, J.-P., 2003. Surface-wave higher mode phase velocity measurements using a roller coaster type algorithm, *Geophys. J. Int.,* **155,** 289–307.

Bézier, P., 1966. Définition numérique des courbes et surfaces I, *Automatisme,* **11,** 625–632.

Bézier, P., 1967. Définition numérique des courbes et surfaces II, *Automatisme,* **12,** 625–632.

Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *J. geophys. Int.,* **178,** 1411–1436.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K. & Rawlinson, N., 2012. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res.,* **117,** B02301, doi:10.1029/2011JB008560.

Boschi, L. & Dziewonski, A.M., 2000. Whole Earth tomography from delay times of P, PcP, PKP phases: lateral heterogeneities in the outer core, or radial anisotropy in the mantle? *J. geophys. Res., 105,* 13 675–13 696.

Boschi, L. & Ekström, G., 2002. New images of the Earth's upper mantle from measurements of surface wave phase velocity anomalies, *J. geophys. Res.,* **107,** doi:10.1029/2000JB000059.

Bozdag, E. & Trampert, J., 2007. On crustal corrections in surface wave tomography, *Geophys. J. Int.,* **172,** 1076–1088.

Cammarano, F. & Romanowicz, B., 2007. Insights into the nature of the transition zone from physically constrained inversion of long-period seismic data, *Proc. Natl. Acad. Sci. U.S.A.,* **104,** 9139–9144.

Cammarano, F., Goes, S., Vacher, P. & Giardini, D., 2003. Inferring upper mantle temperatures from seismic velocities, *Phys. Earth planet. Inter.,* **138,** 197–222.

Cammarano, F., Deuss, A., Goes, S. & Giardini, D., 2005. One dimensional physical reference models for the upper mantle and transition zone: combining seismic and mineral physics constraints, *J. geophys. Res.,* **110,** doi:10.1029/2004JB003272.

Cammarano, F., Romanowicz, B., Stixrude, L., Lithgow-Bertelloni, C. & Xu, W., 2009. Inferring the thermochemical structure of the upper mantle from seismic data, *Geophys. J. Int.,* **179,** 1169–1185.

Cara, M., 1979. Lateral variations of *S*-velocity in the upper mantle from higher Rayleigh modes, *Geophys. J. R. astr. Soc.,* **57,** 649–670.

Cobden, L., Goes, S., Cammarano, F. & Connolly, J., 2008. Thermochemical interpretation of one-dimensional seismic reference models for the upper mantle: evidence for bias due to heterogeneity, *Geophys. J. Int.,* **175,** 627–648.

Cobden, L., Goes, S., Ravenna, M., Styles, E., Cammarano, F., Gallagher, K. & Connolly, J., 2009. Thermochemical interpretation of 1-D seismic data for the lower mantle: the significance of non-adiabatic thermal gradients and compositional heterogeneity, *J. geophys. Res.,* **114,** B11309, doi:10.1029/2008JB006262.

Duffy, T.S., Zha, C., Downs, R.T., Mao, H.K. & Hemley, R.J., 1995. Elastic constants of forsterite Mg_2 SiO_4 to 16 GPa, *Nature,* **378,** 170–173.

Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary Reference Earth Model, *Phys. Earth planet. Inter.,* **25,** 297–356.

Ekström, G. & Dziewonski, A.M., 1998. The unique anisotropy of the Pacific upper mantle, *Nature,* **394,** 168–172.

Ferreira, A., Woodhouse, J., Visser, K. & Trampert, J., 2010. On the robustness of global radially anisotropic surface wave tomography, *J. geophys. Res.,* **115,** B04313, doi:10.1029/2009JB006716.

Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y. & Villasenor, A., 2013. Multiscale full waveform inversion, *Geophys. J. Int.,* **194,** 534–556.

Fukao, Y., Widiyantoro, S. & Obayashi, M., 2001. Stagnant slabs in the upper and lower mantle transition region, *Rev. of Geophysics,* **39,** 291–323.

Gaboret, C., Forte, A.M. & Montagner, J.-P., 2003. The unique dynamics of the Pacific hemisphere mantle and its signature on seismic anisotropy, *Earth planet. Sci. Lett.,* **208,** 219–233.

Ganguly, J., Freed, A.M. & Saxena, S.K., 2009. Density profiles of oceanic slabs and surrounding mantle: integrated thermodynamic and thermal modeling, and implications for the fate of slabs at the 660 km discontinuity, *Phys. Earth planet. Inter.,* **172,** 257–267.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B., 2003. *Bayesian Data Analysis,* Chapman & Hall/CRC, 696, p.

Gilbert, F. & Dziewonski, A.M., 1975. An application of normal mode theory to the retrieval of structural parameters and source mechanisms from seismic spectra, *Phil. Trans. R. Soc. Lond. A.,* **278,** 187–269.

Goes, S., Govers, R. & Vacher, P., 2000. Shallow mantle temperatures under Europe form *P* and *S* wave tomography, *J. geophys. Res.,* **105,** 11 153–11 169.

Green, P., 1995. Reversible jump MCMC computation and Bayesian model selection, *Biometrika,* **82,** 711–732.

Green, P.J. & Mira, A., 2001. Delayed rejection in reversible jump Metropolis-Hastings, *Biometrika,* **88,** 1035–1053.

Gung, Y., Panning, M. & Romanowicz, B., 2003. Global anisotropy and the thickness of continents, *Nature,* **422,** 707–711.

Hashin, Z. & Shtrikman, S., 1963. A variational approach to the theory of the elastic behaviour of multiphase materials, *Mech. Phys. Solids,* **11,** 12–140.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrika,* **57,** 97–109.

Hauser, J., Dyer, K.M., Pasyanos, M.E., Bungum, H., Faleide, J.I., Clark, A. & Schweitzer, J., 2011. A probabilistic seismic model for the European Artic, *J. geophys. Res.,* **116,** B01303, doi:10.1029/2010JB007889.

Hirshmann, M., 2000. Mantle solidus: experimental constraints and the effects of peridotite composition, *Geochem. Geophys. Geosyst.,* **1,** doi:10.1029/2000GC000070.

Hirshmann, M., Tenner, T., Aubaud, C. & Withers, A., 2009. Dehydration melting of nominally anhydrous mantle: the primacy of partitioning, *Phys. Earth planet. Inter.,* **176,** 54–68.

Isaak, D.G., 1992. High-temperature elasticity of iron-bearing olivines, *J. geophys. Res.,* **97,** 1871–1885.

Ita, J. & Stixrude, L., 1992. Petrology, elasticity and composition of the mantle transition zone, *J. geophys. Res.,* **97,** 6849–6866.

Ito, E. & Takahashi, E., 1989. Postspinel transformations in the system Mg_2 SiO_4-Fe_2 SiO_4 and some geophysical implications, *J. geophys. Res.,* **94,** 10 637–10 646.

Jackson, I., 1998. Elasticity, composition and temperature of the Earth's lower mantle: a reappraisal, *Geophys. J. Int.,* **134,** 291–311.

Katsura, T. *et al.*, 2004. Olivine-wadsleyite transition in the system $(Mg,Fe)_2$ SiO_4, *J. geophys. Res.,* **109,** doi:10.1029/2003JB002438.

Katsura, T., Yoneda, A., Yamazaki, D., Yoshino, T. & Ito, E., 2010. Adiabatic temperature profile in the mantle, *Phys. Earth planet. Inter.,* **183,** 212–218.

Kennett, B. L.N., Engdahl, E.R. & Buland, R., 1995. Constraints on seismic velocities in the Earth from travel times, *Geophys. J. Int.,* **122,** 108–124.

Khan, A. & Shankland, T.J., 2012. A geophysical perspective on mantle water content and melting: inverting electromagnetic sounding data using laboratory-based electrical conductivity profiles, *Earth planet. Sci. Lett.,* **317,** 27–43.

Khan, A., Connolly, A.D. & Olsen, A.N., 2006. Constraining the composition and thermal state of the mantle beneath Europe from inversion of long-period electromagnetic sounding data, *J. geophys. Res.,* **111,** B10102, doi:10.1029/2006JB004270.

Khan, A., Boschi, L. & Connolly, A.D., 2009. On mantle chemical and thermal heterogeneities and anisotropy as mapped by inversion of global surface wave data, *J. geophys. Res.,* **114,** B09305, doi:10.1029/2009JB006399.

Khan, A., Boschi, L. & Connolly, J.A.D., 2011. Mapping the Earth's thermochemical and anisotropic structure using global surface wave data, *J. geophys. Res.,* **116,** B01301, doi:10.1029/2010JB007828.

Kustowski, B., Ekström, G. & Dziewonski, A.M., 2008. Anisotropic shear wave velocity structure of the Earth's mantle: a global model, *J. geophys. Res.,* **113,** B06306, doi:10.1029/2007JB005169.

Litasov, K., Ohtani, E., Sano, A., Suzuki, A. & Funakoshi, K., 2005. Wet subduction versus cold subduction, *Geophys. Res. Lett.,* **32,** L13312, doi:10.1029/2005GL022921.

Love, A.E.H., 1927. A Treatise on the Mathematical Theory of Elasticity, Cambridge University Press, 643 p.

MacKay, D.J.C., 2003. *Information Theory, Inference and Learning Algorithms,* Cambridge University Press, 640 p.

Mainprice, D., 2007. Seismic anisotropy of the deep earth from a mineral and rock physics perspective, in *Treatise on Geophysics: Mineral Physics,* pp. 437–491, ed. Price, G.D., Elsevier.

Mégnin, C. & Romanowicz, B., 2000. The three-dimensional shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms, *Geophys. J. Int.,* **143,** 709–728.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E., 1953. Equation of state calculations by fast computing machines, *J. Chem. Phys.,* **21,** 1087–1091.

Montagner, J.-P. & Jobert, N., 1988. Vectorial tomography-II. Application to the Indian ocean, *Geophys. J.,* **94,** 309–344.

Montagner, J.-P. & Kennett, B.L.N., 1996. How to reconcile body-wave and normal-mode reference Earth models, *Geophys. J. Int.,* **125,** 229–248.

Montagner, J.-P. & Tanimoto, T., 1990. Global anisotropy in the upper mantle inferred from the regionalization of phase velocities, *J. geophys. Res.,* **95,** 4797–4819.

Mosca, I., Cobden, L., Deuss, A., Ritsema, J. & Trampert, J., 2012. Seismic and mineralogical structures of the lower mantle from probabilistic tomography, *J. geophys. Res.,* **117,** B06304, doi:10.1029/2011JB008851.

Mosegaard, K., 1998. Resolution analysis of general inverse problems through inverse Monte-Carlo sampling, *Inverse Problems,* **14,** 405–426.

Mosegaard, K. & Tarantola, A., 1995. Monte-Carlo sampling of solutions to inverse problems, *J. geophys. Res.,* **100,** 12 431–12 447.

Nettles, M. & Dziewonski, A.M., 2008. Radially anisotropic shear-velocity structure of the upper mantle globally and beneath North America, *J. geophys. Res.,* **113,** B02303, doi: 10.1029/2006JB004819.

Panning, M. & Romanowicz, B., 2006. A three-dimensional radially anisotropic model of shear velocity in the whole mantle, *Geophys. J. Int.,* **167,** 361–379.

Panning, M., Lekic, V. & Romanowicz, B., 2010. The importance of crustal corrections in the development of a new global model of radial anisotropy, *J. geophys. Res.,* **115,** doi:10.1029/2010JB007520.

Ringwood, A.E., 1975. *Composition and Petrology of the Earth's Mantle,* McGraw-Hill, 618 p.

Ritsema, J., van Heijst, H.J. & Woodhouse, J.H., 1999. Complex shear wave velocity structure imaged beneath Africa and Iceland, *Science,* **286,** 1925–1928.

Romanowicz, B., 2003. Global mantle tomography: progress status in the last 10 years, *Ann. Rev. Earth planet. Sci.,* **31,** 303–328.

Rosenthal, J., 2000. Parallel computing and Monte Carlo algorithms, *Far Est J. Theor. Statist.,* **4,** 207–236.

Shapiro, N.M. & Ritzwoller, M.H., 2002. Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle, *Geophys. J. Int.,* **151,** 88–105.

Shearer, P.M., 2000. Upper mantle seismic discontinuities, in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale,* pp. 115–131, eds Karato, S.I., Forte, A.M., Liebermann, R.C., Masters, G. & Stixtude, L., American Geophysical Union.

Shen, W., Ritzwoller, M. & Schlute-Pelkum, V., 2013a. A 3-D model of the crust and uppermost mantle beneath the Central and Western US by joint inversion of receiver functions and surface wave dispersion, *J. geophys. Res.,* **118,** 262–276.

Shen, W., Ritzwoller, M., Schlute-Pelkum, V. & Lin, F., 2013b. Joint inversion of surface wave dispersion and receiver functions: a Bayesian Monte-Carlo approach, *Geophys. J. Int.,* **192,** 807–836.

Stixrude, L., 1997. Structure and sharpness of phase transitions and mantle discontinuities, *J. geophys. Res.,* **102,** 14 835–14 582.

Stixrude, L. & Lithgow-Bertelloni, C., 2010. Thermodynamics of the Earth's mantle, *Rev. Min. Geoch.,* **71,** 465–484.

Tagawa, M., Nakakuki, T. & Tajima, F., 2007. Dynamical modeling of trench retreat driven by the slab interaction with the mantle transition zone, *Earth Planets Space,* **59,** 65–74.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* Society for Industrial and Applied Mathematics, 342 p.

Trampert, J. & Woodhouse, J.H., 2003. Global anisotropic phase velocity maps for fundamental mode surface waves between 40 and 150 s, *Geophys. J. Int.,* **154,** 154–165.

Trampert, J., Deschamps, F., Resovsky, J. & Yuen, D., 2004. Probabilistic tomography maps chemical heterogeneities throughout the lower mantle, *Science,* **306,** 853–856.

Vacher, P., Mocquet, A. & Sotin, C., 1996. Comparison between tomographic structures and models of convection in the upper mantle, *Geophys. J. Int.,* **124,** 45–56.

Vacher, P., Mocquet, A. & Sotin, C., 1998. Computation of seismic profiles from mineral physics: the importance of the non-olivine components for explaining the 660 km depth discontinuity, *Phys. Earth planet. Inter.,* **106,** 275–298.

van der Hilst, R.D., Widyantoro, S. & Engdahl, E.R., 1997. Evidence for deep mantle circulation from global tomography, *Nature,* **386,** 578–584.

van Heijst, H.J. & Woodhouse, J.H., 1997. Measuring surface-wave overtone phase velocities using a mode-branch stripping technique, *Geophys. J. Int.,* **131,** 209–230.

Verhoeven, O. *et al.*, 2005. Interior structure of terrestrial planets: modeling Mars' mantle and its electromagnetic, geodetic, and seismic properties, *J. geophys. Res.,* **110,** doi:10.1029/2008JB005678.

Verhoeven, O. *et al.*, 2009. Constraints on thermal state and composition of the Earth's lower mantle from electromagnetic impedances and seismic data, *J. geophys. Res.,* **114,** doi:10.1029/2004JE002271.

Visser, K., Lebedev, S., Trampert, J. & Kennett, B. L.N., 2007. Global Love wave overtone measurements, *Geophys. Res. Lett.,* **34,** L03302, doi:10.1029/2006GL028671.

Visser, K., Trampert, J. & Kennett, B.L.N., 2008a. Global anisotropic phase-velocity maps for higher mode Love and Rayleigh waves, *Geophys. J. Int.,* **172,** 1016–1032.

Visser, K., Trampert, J., Lebedev, S. & Kennett, B.L.N., 2008b. Probability of radial anisotropy in the deep mantle, *Earth planet. Sci. Lett.,* **270,** 241–250.

Watt, J.P., Davies, G.F. & O'Connell, R.J., 1976. The elastic properties of composite materials, *Rev. geophys. Space Phys.,* **14,** 541–563.

Wessel, P. & Smith, W.H.F., 1998. New, improved version of the generic mapping tools released, *EOS, Trans. Am. geophys. Un.,* **79,** 579.

Woodhouse, J., 1988. The calculation of eigenfrequencies and eigenfunctions of the free oscillations of the Earth and the Sun, in *Seismological Algorithms,* pp. 321–370, ed. Doornbos, D., Academic Press.

Zhu, H., Bozdag, E., Peter, D. & Tromp, J., 2012. Structure of the European upper mantle revealed by adjoint tomography, *Nature Geosci.,* **5,** 493–498.